

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: May 16, 2003, 15:44:37 ; Search time 401.589 Seconds  
(without alignments)  
1449.381 Million cell updates/sec

Title: US-10-052-092-15

Perfect score: 20

Sequence: 1 caagcgccagagagatgatg 20

Scoring table: IDENTITY\_NTC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

GenBank:\*

1: gb\_ha:\*

2: gb\_htg:\*

3: gb\_in:\*

4: gb\_cm:\*

5: gb\_ov:\*

6: gb\_pat:\*

7: gb\_ph:\*

8: gb\_pl:\*

9: gb\_pr:\*

10: gb\_ro:\*

11: gb\_sts:\*

12: gb\_sy:\*

13: gb\_un:\*

14: gb\_vl:\*

15: em\_ba:\*

16: em\_fun:\*

17: em\_hum:\*

18: em\_in:\*

19: em\_mu:\*

20: em\_om:\*

21: em\_or:\*

22: em\_ov:\*

23: em\_pat:\*

24: em\_ph:\*

25: em\_pl:\*

26: em\_ro:\*

27: em\_sts:\*

28: em\_un:\*

29: em\_vl:\*

30: em\_htg\_hum:\*

31: em\_htg\_in:\*

32: em\_htg\_other:\*

33: em\_htg\_mus:\*

34: em\_htg\_pln:\*

35: em\_htg\_rtd:\*

36: em\_htg\_man:\*

37: em\_htg\_vtl:\*

38: em\_sy:\*

39: em\_htgo\_hum:\*

40: em\_htgo\_mus:\*

41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	26	E13443	E13443 PCR primer
2	20	100.0	144	S80316	S80316 Oestrogen r
3	20	100.0	392	AF120105	AF120105 Homo sapi
4	20	100.0	521	AX004781	AX004781 Sequence
5	20	100.0	539	AX004793	AX004793 Sequence
6	20	100.0	566	AX004790	AX004790 Sequence
7	20	100.0	1223	AX066401	AX066401 Sequence
8	20	100.0	1237	AX066402	AX066402 Sequence
9	20	100.0	1374	AX066401	AX066401 Sequence
10	20	100.0	1788	SSSTREC	SSSTREC
11	20	100.0	1788	AXA11751	AXA11751 Sequence
12	20	100.0	1966	OAESTRRC	OAESTRRC
13	20	100.0	2092	AX474709	AX474709 Sequence
14	20	100.0	2092	108538	108538 Sequence 1
15	20	100.0	2092	HUMRMCF	HUMRMCF
16	20	100.0	2106	BSU47678	BSU47678
17	20	100.0	3222	A42099	A42099 Sequence 1
18	20	100.0	3746	AF124093	AF124093 Equus cab
19	20	100.0	4963	AR029418	AR029418 Sequence
20	20	100.0	4963	115368	115368 Sequence 16
21	20	100.0	4963	156760	156760 Sequence 9
22	20	100.0	5653	AY033393	AY033393 Ovis arlie
23	20	100.0	6450	ARI53585	ARI53585 Sequence
24	20	100.0	6450	HSERR	HSERR
25	20	100.0	6610	AF061181	AF061181 Mamalian
26	20	100.0	6639	AX128351	AX128351 Sequence
27	20	100.0	6695	AX128347	AX128347 Sequence
28	20	100.0	6695	AX128353	AX128353 Sequence
29	20	100.0	6695	AX128354	AX128354 Sequence
30	20	100.0	6801	AX128355	AX128355 Sequence
31	20	100.0	6818	AX128346	AX128346 Sequence
32	20	100.0	6833	AX128349	AX128349 Sequence
33	20	100.0	6900	AX128341	AX128341 Sequence
34	20	100.0	6956	AX128348	AX128348 Sequence
35	20	100.0	7038	AX128342	AX128342 Sequence
36	20	100.0	139687	AL590993	AL590993 Human DNA
37	20	100.0	165237	AX232505	AX232505 Sequence
38	20	100.0	349980	AX232503	AX232503 Sequence
39	20	100.0	349980	AX453703	AX453703 Sequence
40	18.4	92.0	748	CPA313195	CPA313195 Canis fam
41	18	90.0	200000	AC007355	AC007355 Homo sapi
42	18	90.0	243358	AC110814	AC110814 Homo sapi
43	17.4	87.0	96599	AC125306	AC125306 Rattus no
44	17.4	87.0	146067	OSJN00150	OSJN00150 Oryza sat
45	17.4	87.0	160889	AC109098	AC109098 Rattus no

## ALIGNMENTS

RESULT 1  
E13443/c  
LOCUS E13443 26 bp DNA linear PAT 27-APR-1998  
DEFINITION PCR primer for detecting mRNA which encode human estrogen receptor.  
ACCESSION E13443  
VERSION E13443.1 GI:3252248  
KEYWORDS JP 1997187299-A/5.  
SOURCE JP 1997187299-A/5.  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 26)  
AUTHORS KIMOTO Y.  
TITLE PRIMER FOR PCR  
JOURNAL Patent: JP 1997187299-A 5 22-JUL-1997;

```

COMMENT
OS NIPPON BIO SERAPIT KK
OC Artificial sequences.
PN JP 1997187299-A/5
PD 22-001-1997
PF 05-JAN-1996 JP 1996027222
PI KIMOTO YASUHIKO
PC C1201/68, C07H21/04, C12N15/09;
CC strandedness: Single;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: Yes;
FH Key
FH Location/Qualifiers
FEATURES
source 1. .26
FT /organism='Artificial sequences' FT
FT misc.feature 1. .26
FT /note='PCR primer E-7'.
FT Location/Qualifiers
1. .26
/organism='unidentified'
/db_xref='taxon:32644'
2 a 8 c 5 g 11 t
BASE COUNT
ORIGIN
Query Match 100.0%; Score 20; DB 6; Length 26;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAAGCGCCAGAGAGATGATG 20
Db 20 CAAGCGCCAGAGATGATG 1
|||||
LOCUS S80316 144 bp mRNA linear PRI 02-APR-1996
DEFINITION oestrogen receptor [alternatively spliced, exons 4 and 7] [human,
ACCESSION S80316
VERSION S80316
KEYWORDS S80316.1 GI:1246086
SOURCE Homo sapiens endometrial liver.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 144)
AUTHORS Daffada A.A. and Dowsett M.
TITLE Tissue-dependent expression of a novel splice variant of the human
oestrogen receptor
JOURNAL J. Steroid Biochem. Mol. Biol. 55 (3-4), 413-421 (1995)
MEDLINE 96132694
PUBMED 8541239
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI gbbasg 173922] from the original journal article.
This sequence comes from Fig. 3.
Location/Qualifiers
1. .144
/organism='Homo sapiens'
/db_xref='taxon:9606'
1. .144
/partial
/feature='oestrogen receptor, delta 4/7 ER'
1. .135
/partial
/feature='oestrogen receptor, delta 4/7 ER'
/partial
/feature='oestrogen receptor, delta 4/7 ER'
/note='This sequence comes from Fig. 3; delta 4/7 ER'
/codon_start=1
/product='oestrogen receptor'
/protein_id='BAB55900.1'
/db_xref='GI:1246087'
/translacion='XHKRQDDEGPPDDGCGRPDPAAMAAPAGPAPPIPHQATHE

```

BASE COUNT	38 a	48 c	40 g	18 t
ORIGIN				
Query Match		100.0%;	Score 20;	DB 9; Length 144;
Best Local Similarity		100.0%;	Pred. No. 6;	
Matches 20; Conservative 0;		Mismatches 0;	Indels 0;	Gaps 0;
QY 1 CAAGCGCCAGAGATGATG 20				
Db 6 CAAGCGCCAGAGATGATG 25				
RESULT 3				
LOCUS AF120105		392 bp	mRNA	linear PRI 06-APR-1999
DEFINITION Homo sapiens alternatively-spliced estrogen receptor alpha mRNA,				
partial cds.				
ACCESSION AF120105				
VERSION AF120105.1		GI:4567039		
KEYWORDS				
SOURCE				
ORGANISM Homo sapiens.				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE 1 (bases 1 to 392)				
AUTHORS Campbell-Thompson,K.L.				
TITLE Direct Submission				
JOURNAL Submitted (13-JAN-1999) Medicine, University of Florida, Box 1002144				
Gastroenterology, Gainesville, FL 32667, USA				
location/Qualifiers				
1..392				
/organism="Homo sapiens"				
/db_xref="taxon:9606"				
/chromosome="6"				
/map="6q25.1"				
/sex="male"				
/cell_type="colon adenocarcinoma"				
<1..343				
/note="steroid hormone receptor; missing exon 5"				
/codon_start=2				
/product="alternatively-spliced estrogen receptor alpha"				
/protein_id="AAD23565.1"				
/db_xref="GI:4567040"				
/translation="RRGRMLKKRRQDGGEGEEVSSAGDMRANMPSPLTKRSK				
KNSLISLTADQWVSALIDAEPIIYSIEDYPRPSPASMGILFTNLADRELYHIMWV				
AKVVGTEKRENV"				
25				
/note="compared to wild type sequence deposited in GenBank				
Accession Number X03635"				
/replace="a"				
variation				
BASE COUNT 97 a 92 c 124 g 79 t				
ORIGIN				
Query Match		100.0%;	Score 20;	DB 9; Length 392;
Best Local Similarity 100.0%;		Pred. No. 6.3;		
Matches 20; Conservative 0;		Mismatches 0;	Indels 0;	Gaps 0;
QY 1 CAAGCGCCAGAGATGATG 20				
Db 28 CAAGCGCCAGAGATGATG 47				
RESULT 4				
LOCUS AX004781		521 bp	DNA	linear PAT 24-AUG-2000
DEFINITION Sequence 1 from Patent W09911785.				
ACCESSION AX004781				
VERSION AX004781.1		GI:9928194		
KEYWORDS				
SOURCE unidentified.				
ORGANISM unidentified.				
REFERENCE 1 (bases 1 to 521)				

AUTHORS Mcgregor,D.  
TITLE Chimeric binding peptide library screening method  
JOURNAL Patent: WO 9911785-A 11-MAR-1999;  
MCGREGOR DUNCAN (GB); ROWETT RESEARCH SERVICES LIMIT (GB)  
FEATURES  
SOURCE 1..521  
/organism="unidentified"  
/db\_xref="taxon:32644"  
/note="Recombinant human oestrogen"  
CDS 41..124  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC04912.1"  
/db\_xref="GI:9928195"  
/translation="MKTLPTAAAGILLAAQPMARVQLQ"  
BASE COUNT 153 a 109 c 153 g 104 t  
ORIGIN  
Query Match 100.0%; Score 20; DB 6; Length 521;  
Best Local Similarity 100.0%; Pred. No. 6.4;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 CAAGCGCCAGAGAGATGATG 20  
|||||  
DB 427 CAAGCGCCAGAGAGATGATG 446  
RESULT 5  
AX004793 539 bp DNA linear PAT 24-AUG-2000  
LOCUS Sequence 13 from Patent WO9911785.  
DEFINITION AX004793  
ACCESSION AX004793.1 GI:9928204  
VERSION  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 539)  
AUTHORS Mcgregor,D.  
TITLE Chimeric binding peptide library screening method  
JOURNAL Patent: WO 9911785-A 11-MAR-1999;  
MCGREGOR DUNCAN (GB); ROWETT RESEARCH SERVICES LIMIT (GB)  
FEATURES  
SOURCE 1..539  
/organism="unidentified"  
/db\_xref="taxon:32644"  
/note="Recombinant human oestrogen"  
CDS 41..481  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC04917.1"  
/db\_xref="GI:9928205"  
/translation="MKTLPTAAAGILLAAQPMARVSAKETRYCANCNDIYASGIH  
YGVMSCEGCKAFPKRSIQGHNDYCPAINQCTIDNRKSCQACRLKCYEVGMKGG  
IKKDRGGRMKHKRQDDEGRGEGSGGSSAQPALIDIAAA"  
BASE COUNT 148 a 118 c 165 g 108 t  
ORIGIN  
Query Match 100.0%; Score 20; DB 6; Length 539;  
Best Local Similarity 100.0%; Pred. No. 6.4;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 CAAGCGCCAGAGAGATGATG 20  
|||||  
DB 385 CAAGCGCCAGAGAGATGATG 404  
RESULT 6  
AX004790 566 bp DNA linear PAT 24-AUG-2000  
LOCUS Sequence 10 from Patent WO9911785.  
DEFINITION AX004790  
ACCESSION AX004790.1 GI:9928202  
VERSION  
SOURCE

KEYWORDS  
SOURCE unidentified.  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 566)  
AUTHORS Mcgregor,D.  
TITLE Chimeric binding peptide library screening method  
JOURNAL Patent: WO 9911785-A 10 11-MAR-1999;  
MCGREGOR DUNCAN (GB); ROWETT RESEARCH SERVICES LIMIT (GB)  
FEATURES  
SOURCE 1..566  
/organism="unidentified"  
/db\_xref="taxon:32644"  
/note="Recombinant human oestrogen"  
CDS 41..124  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC04916.1"  
/db\_xref="GI:9928203"  
/translation="MKTLPTAAAGILLAAQPMARVQLQ"  
BASE COUNT 164 a 118 c 169 g 115 t  
ORIGIN  
Query Match 100.0%; Score 20; DB 6; Length 566;  
Best Local Similarity 100.0%; Pred. No. 6.4;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 CAAGCGCCAGAGAGATGATG 20  
|||||  
DB 427 CAAGCGCCAGAGAGATGATG 446  
RESULT 7  
AX066401 1223 bp DNA linear PAT 24-JAN-2001  
LOCUS Sequence 1 from Patent WO0100823.  
DEFINITION AX066401  
ACCESSION AX066401.1 GI:12544111  
VERSION  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 1223)  
AUTHORS Gannon,F., Denger,S. and Flouriot,G.  
TITLE Novel isoforms of the human estrogen receptor -g(a)  
JOURNAL Patent: WO 0100823-A 1 04-JAN-2001;  
EUROPEAN MOLECULAR BIOLOGY LABORATORY (DE)  
FEATURES  
SOURCE 1..1223  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 303 a 316 c 341 g 263 t  
ORIGIN  
Query Match 100.0%; Score 20; DB 6; Length 1223;  
Best Local Similarity 100.0%; Pred. No. 6.7;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 CAAGCGCCAGAGAGATGATG 20  
|||||  
DB 236 CAAGCGCCAGAGAGATGATG 255  
RESULT 8  
HSESRL3 1237 bp DNA linear PRI 07-APR-2000  
LOCUS Homo sapiens estrogen receptor alpha (ESR1) gene, exon 4.  
DEFINITION HSESRL3  
ACCESSION AF123496  
VERSION AF123496.1 GI:5821720  
KEYWORDS  
SEGMENT 3 of 7  
SOURCE Homo sapiens.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1237)  
AUTHORS Schbert, E.L., Lee, M.K., Newman, B. and King, M.C.  
TITLE Single nucleotide polymorphisms (SNPs) in the estrogen receptor  
gene and breast cancer susceptibility  
JOURNAL Steroid Biochem. Mol. Biol. 71 (1-2), 21-27 (1999)  
MEDLINE 20084372  
PUBMED 10619354  
REFERENCE 2 (bases 1 to 1237)  
AUTHORS Schbert, E.L., Lee, M.K. and King, M.C.  
TITLE Direct Submission  
JOURNAL Submitted (26-JAN-1999) Medical Genetics, University of Washington,  
1959 NE Pacific Street, Seattle, WA 98195-7720, USA  
FEATURES  
Source  
1. 1237  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/map="6q25"  
/gene="ESR1"  
/number="4"  
exon 350..685  
BASE COUNT 369 a 218 c 286 g 364 t  
ORIGIN  
Query Match 100.0%; Score 20; DB 9; Length 1237;  
Best Local Similarity 100.0%; Pred. No. 6.7;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CY 1 CAAGCCGACAGAGATGATG 20  
DB 390 CAAGCCGACAGAGATGATG 409  
RESULT 9  
LOCUS AX066402 1374 bp DNA linear PAT 24-JAN-2001  
DEFINITION Sequence 2 from Patent WO0100823.  
ACCESSION AX066402  
VERSION AX066402.1 GI:12544112  
KEYWORDS human.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1374)  
AUTHORS Gannon, F., Denger, S. and Flouriot, G.  
TITLE Novel isoforms of the human estrogen receptor- $\alpha$  (a)  
JOURNAL Patent: WO 0100823-A 2 04-JAN-2001;  
EUROPEAN MOLECULAR BIOLOGY LABORATORY (DE)  
FEATURES  
Source  
1. 1374  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 349 a 346 c 386 g 293 t  
ORIGIN  
Query Match 100.0%; Score 20; DB 6; Length 1374;  
Best Local Similarity 100.0%; Pred. No. 6.7;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CY 1 CAAGCCGACAGAGATGATG 20  
DB 387 CAAGCCGACAGAGATGATG 406  
RESULT 10  
LOCUS S5ESTREC 1788 bp mRNA linear MAM 15-SEP-1994  
DEFINITION S.scrofa mRNA for estradiol receptor.  
ACCESSION 237167

VERSION 237167.1 GI:587554  
KEYWORDS estradiol receptor.  
SOURCE Sus scrofa.  
ORGANISM Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.  
REFERENCE 1 (bases 1 to 1788)  
AUTHORS Bokenkamp, D., Jungblut, P.W. and Thole, H.H.  
TITLE The C-terminal half of the porcine estradiol receptor contains no  
post-translational modification: determination of the primary  
structure  
JOURNAL Mol. Cell. Endocrinol. 104 (2), 163-172 (1994)  
MEDLINE 95080454  
PUBMED 7988744  
REFERENCE 2 (bases 1 to 1788)  
AUTHORS Bokenkamp, D.  
TITLE Direct Submission  
JOURNAL Submitted (07-SEP-1994) Dirk Bokenkamp, Max-Planck-Institut fuer  
experimentelle Endokrinologie, Feodor-Lynen Strasse 7, Hannover,  
30625, Germany  
FEATURES  
Source  
1. 1788  
/organism="Sus scrofa"  
/db\_xref="taxon:9823"  
/sex="Female"  
/tissue\_type="uterus"  
1. 1788  
/codon\_start=1  
/product="estradiol receptor"  
/protein\_id="CA85524.1"  
/db\_xref="GI:587555"  
/db\_xref="SWISS-PROT:O29040"  
/translation="MFMTHKRSAGMALLQIQIANLEPLNPKIPLERIGEVY  
DSKRPATYNTPEGAIDFRMAAASAVYQSGSLAYGPSESAATGANGLGFPPLNS  
VSPSPVLPLAPPQSLPLHPGQGVYVLENEPSSGVAVREAGPAFVPSNDRRG  
GRIELASTSDKSGNAMESAKERTCAVCNDYASGYARGVASCCKAFKRSIGND  
YMCPTAUNOCTIDKNRKSQCAQLKCYEVGAKGRIEDRGRGKYLKKRORDGSG  
RNEVPPGDRSANLAMPSPILIKHKNSVLSLFDQVHISALLEAPETIYSRDPY  
RPSLSASMGILLTNLADREIVHIMWAKRVPGLDLSHDQVHILKCAWELIMIGYV  
WRSEHFGKILRAPNILLDRNGCYGSAWEITDMILATSSFRMANIOGEPYCKLS  
ITLNSGVITFLSTSLERDHIHVLDKITYDTLHIMARAAGTLQOQRRRLAQLL  
LILSHFHMNSKGMELYNMCKNVPLDYDLLEMDARHLRAPNIGGPPEDMSGS  
QLATSGSTPESHSLQWYITIGEAENFPYTI"  
BASE COUNT 410 a 530 c 504 g 344 t  
ORIGIN  
Query Match 100.0%; Score 20; DB 4; Length 1788;  
Best Local Similarity 100.0%; Pred. No. 6.8;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CY 1 CAAGCCGACAGAGATGATG 20  
DB 801 CAAGCCGACAGAGATGATG 820  
RESULT 11  
LOCUS AX411751 1788 bp DNA linear PAT 14-JUN-2002  
DEFINITION Sequence 1 from Patent WO0228175.  
ACCESSION AX411751  
VERSION AX411751.1 GI:21444272  
KEYWORDS human.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Chambon, P. and Metzger, D.  
TITLE Transgenic mouse for targeted recombination mediated by modified  
Cre-er  
JOURNAL Patent: WO 0228175-A 1 11-APR-2002;  
ASS POOR DE DEV DE LA RECH (FR)  
FEATURES  
Location/Qualifiers

```

source
1..1788
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..1788
CDS
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD35258.1"
/db_xref="gi:21444273"
/translation="MTWTLITKSGMALLHIOGNELEPIINRPOKILPERDLEEVY
DSKPAVNYPEGAAYEPNAAAAAAYGOTGCLPYRGSAAAPGNSGIGFEPINS
VSPSPMLHPROPOLSPLOPHGOVPIYINNESGYTVAEAPPAFYRPNNDNRG
GGERLASTNDGSKAMBSAKETRICVCHNDYASGHTGVSCGCAFKRSIQGND
TMCPTNQCITDKNRKSCQACLRKCYEYGMKGRKDKRGRLKHKRQDGE
RGEVSGADMRAANIPSPMLTKRSKSLSLTADQYMSALLDAEPIILYSEDP
RPESSAMGILLTADRELPHIMAKRYGEVDLTLEOVHLCAMLETIMGIV
WRSMPEKLLFAPNILLDRNGKCYEYGMKGRKDKRGRLKHKRQDGE
IILNSGYRPLSTLKSLEERDHRHVDITITPILHAKAGDULOORHLMQEL
LILSHIRMSKGMKHEHLSKCNVPPYLDLLEMLHRLHAPTSRGASVEETDSS
HLATGSSSHSLQKTYTGEAEPVAV"

BASE COUNT      418 a      515 c      508 g      347 t
ORIGIN
Query Match      100.0%; Score 20; DB 6; Length 1788;
Best Local Similarity 100.0%; Pred. No. 6.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CAAGCGCCAGAGAGATGATG 20
|||||
DB      801 CAAGCGCCAGAGAGATGATG 820

RESULT 12
OAE5REC      1966 bp      mRNA      linear      MAM 08-OCT-1996
LOCUS      OAE5REC
DEFINITION      O.aries mRNA for estrogen receptor.
ACCESSION      Z49257
VERSION      Z49257.1 GI:1617201
KEYWORDS      estrogen receptor.
SOURCE      Ovis aries.
ORGANISM      Ovis aries
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Caprinae; Ovis.
1 (bases 1 to 1966)
Madigou,T., Tiffoche,C., Lazennec,G., Pelletier,J. and
Thieulant,M.L.
The sheep estrogen receptor: cloning and regulation of expression
in the Hypothalamo-pituitary axis
Mol. Cell. Endocrinol. 121 (2), 153-163 (1996)
JOURNAL
MEDLINE      97047396
PUBMED      8892316
REFERENCE      2 (bases 1 to 1966)
AUTHORS      Madigou,T.
TITLE      Direct Submision
JOURNAL      Submitted (12-MAY-1995) Thierry Madigou, Interactions Cellulaires
et Moleculaires, CNRS URA 256, Rennes, 35042, FRANCE
FEATURES
SOURCE      Location/Qualifiers
1..1966
/organism="Ovis aries"
/strain="Romanov"
/db_xref="taxon:9940"
/clone="lambda p112"
/sex="female"
/tissue_type="anterior pituitary gland"
/clone_lib="cDNA library"
/dev_stage="adult"
1..1966
/product="estrogen receptor"

BASE COUNT      439 a      604 c      549 g      374 t
ORIGIN
Query Match      100.0%; Score 20; DB 4; Length 1966;
Best Local Similarity 100.0%; Pred. No. 6.9;

```

```

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CAAGCGCCAGAGAGATGATG 20
|||||
DB      953 CAAGCGCCAGAGAGATGATG 972

RESULT 13
AX474709      2092 bp      DNA      linear      PAT 12-AUG-2002
LOCUS      AX474709
DEFINITION      Sequence 9 from Patent WO0222882.
ACCESSION      AX474709
VERSION      AX474709.1 GI:22214043
KEYWORDS
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE      1
AUTHORS      Stanton,M., Epstein,D. and Hamaguchi,N.
TITLE      Target activated nucleic acid biosensor and methods of using same
JOURNAL      Patent: WO 0222882-A 9 21-MAR-2002;
Archemix Corporation (US)
FEATURES
SOURCE      Location/Qualifiers
1..2092
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT      473 a      605 c      593 g      421 t
ORIGIN
Query Match      100.0%; Score 20; DB 6; Length 2092;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CAAGCGCCAGAGAGATGATG 20
|||||
DB      1093 CAAGCGCCAGAGAGATGATG 1112

RESULT 14
I08538      2092 bp      DNA      linear      PAT 02-DEC-1994
LOCUS      I08538
DEFINITION      Sequence 1 from Patent WO 8705049.
ACCESSION      I08538
VERSION      I08538.1 GI:588753
KEYWORDS
SOURCE      Unknown.
ORGANISM      Unknown.
REFERENCE      1 (bases 1 to 2092)
AUTHORS      Shine,J.
TITLE      EDCARTOTIC EXPRESSION OF STEROID RECEPTOR PROTEINS
JOURNAL      Patent: WO 8705049-A 1 27-AUG-1987;
FEATURES
SOURCE      Location/Qualifiers
1..2092
/organism="unknown"

BASE COUNT      473 a      605 c      593 g      421 t
ORIGIN
Query Match      100.0%; Score 20; DB 6; Length 2092;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CAAGCGCCAGAGAGATGATG 20
|||||
DB      1093 CAAGCGCCAGAGAGATGATG 1112

RESULT 15
H096RCF      2092 bp      mRNA      linear      PRI 08-NOV-1994
LOCUS      H096RCF
DEFINITION      Human estrogen receptor mRNA, complete cds.
ACCESSION      M12674

```

VERSION M12674.1 GI:182192  
KEYWORDS estrogen receptor.  
SOURCE Human breast cancer cell line MCF-7, cDNA to mRNA, clone OR8.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2092)  
AUTHORS Greene,G.L., Gilna,P., Waterfield,M., Baker,A., Hort,Y. and  
Shine,U.  
TITLE Sequence and expression of human estrogen receptor complementary  
DNA  
JOURNAL Science 231 (4742), 1150-1154 (1986)  
MEDLINE 86122927  
PubMed 3753802  
FEATURES  
SOURCE Location/Qualifiers  
1..2092  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="6q24-q27"  
/cell\_line="MCF-7"  
/cell\_type="Mammary Carcinoma"  
/lab\_host="Prokaryotae"  
/tissue\_lib="lambda gt11"  
1..2092  
/gene="ESR"  
<1..>2092  
/gene="ESR"  
/product="ER mRNA"  
293..2080  
/note="estrogen receptor"  
/codon\_start=1  
/protein\_id="AA52399.1"  
/db\_xref="GI:182193"  
/db\_xref="GDB:600-119-120"  
/translation="MTMTLHTKASGMALHQLQGNLEPLNRPQLPIERPLGEVYL  
DSKPAYNTPGCAAYEFNAANAAYGGTGLPYGSEAAAGSNGLGPPPLNS  
VSPSPMLAPPOLSPFIOPHOVPPYLENEPSGYVREAGPPATFRPNSDRROG  
GREPLASTNDKSGMAESAKETRYCAVNDYASVHYHGWSCGCKAFKRSIOGND  
TMCPTNQCITIDNRKSCQACRLKCEYGMKGGIRKDRSGRLKHKRQDDSGS  
RGEVGSAGDRAANLNPSPMLTKRSKNSALSITADQVSALDAEPILYSEYDPT  
RPFSSASMGILTNLADRELVMINAKRVPGFVLDLHDQVHLFCAMLEIIMIGLY  
WRSMHEPVKLLFAPNLLDRNQKCEGMEVIEFMDLATSRRMMNLQGEFFCLKS  
IILNSGYVTFSLKSEKDHRYLKDITDPLIHMAKAGLTLQOQHORLAOLL  
LILSHIRMSNKGMEHLYMKCKNVPLYLLEMLDAHRLHAPTSRGSASVEETDGS  
HIAIRGSTSSHSLOKTYITGEXGPPATV"

BASE COUNT 473 a 605 c 593 g 421 t  
ORIGIN 1 bp upstream of EcoRI site.

Query Match 100.0%; Score 20; DB 9; Length 2092;  
Best Local Similarity 100.0%; Pred. No. 6.9;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGCGCCAGAGATGATG 20  
|||||  
DB 1093 CAAGCGCCAGAGATG 1112

Search completed: May 17, 2003, 00:56:23  
Job time : 406.369 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 16, 2003, 12:48:16 ; Search time 111.126 Seconds

(Without alignments)  
405,306 Million cell updates/sec

Title: US-10-052-092-15

Perfect score: 20

Sequence: 1 caagcgccagagagatgatg 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

N.Geneseq\_101002:\*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	20	100.0	26	18	AAV00167
2	20	100.0	521	20	AAV00167
3	20	100.0	539	20	AAV00167
4	20	100.0	566	20	AAV00167
5	20	100.0	1223	22	AAV00167
6	20	100.0	1374	22	AAV00167
7	20	100.0	1380	22	AAV00167
8	20	100.0	1788	24	AAV00167
9	20	100.0	2092	8	AAV00167

10	20	100.0	2092	22	AAV00167	Nucleotide sequenc
11	20	100.0	2220	24	AAV00167	Human oestrogen re
12	20	100.0	2322	24	AAV00167	Fused ER protein c
13	20	100.0	2322	24	AAV00167	SSR-LBD fusion pol
14	20	100.0	4963	16	AAV00167	PhCMV*-1 promoter
15	20	100.0	4963	16	AAV00167	PhCMV*-1-controlle
16	20	100.0	4963	20	AAV00167	CDNA encoding rabb
17	20	100.0	4963	20	AAV00167	Rabbit progesteron
18	20	100.0	4963	20	AAV00167	PhCMV*-1 rabbit pr
19	20	100.0	4963	22	AAV00167	Rabbit progesteron
20	20	100.0	6450	20	AAV00167	Human estrogen rec
21	20	100.0	6450	22	AAV00167	Plasmaid E2CDBS e
22	20	100.0	6695	22	AAV00167	Plasmaid C7IBBDS e
23	20	100.0	6695	22	AAV00167	Plasmaid C7IBBDS e
24	20	100.0	6695	22	AAV00167	Plasmaid C7IBBDS e
25	20	100.0	6801	22	AAV00167	Plasmaid C7IBBDS e
26	20	100.0	6818	22	AAV00167	Plasmaid C7IBBDS e
27	20	100.0	6833	22	AAV00167	Plasmaid C7IBBDS e
28	20	100.0	6900	22	AAV00167	Plasmaid C7IBBDS e
29	20	100.0	6956	22	AAV00167	Plasmaid C7IBBDS e
30	20	100.0	7038	22	AAV00167	Directed evolution
31	20	100.0	11365	24	AAV00167	Human oestrogen re
32	20	100.0	465237	24	AAV00167	Human oestrogen re
33	20	100.0	465237	24	AAV00167	Human oestrogen re
34	20	100.0	465237	24	AAV00167	Human oestrogen re
35	18	90.0	550	20	AAV00167	Human ORF ORF403
36	17.4	87.0	288	21	AAV00167	Rat hepatocyte car
37	17.4	87.0	431	21	AAV00167	Aspergillus niger
38	16.4	82.0	564	22	AAV00167	Human breast cance
39	16.4	82.0	568	22	AAV00167	Human breast cance
40	16.4	82.0	1047	24	AAV00167	Endocrine disruptl
41	16.4	82.0	1470	22	AAV00167	Angiotensin conver
42	16.4	82.0	1630	22	AAV00167	Angiotensin conver
43	16.4	82.0	2223	23	AAV00167	Drosophila melanog
44	16.4	82.0	5088	23	AAV00167	Drosophila melanog
45	16.4	82.0	13837	23	AAV00167	Drosophila melanog

#### ALIGNMENTS

RESULT 1	AAV00167/c	AAV00167 standard; DNA; 26 BP.
ID	AAV00167	standard; DNA; 26 BP.
AC	AAV00167	
DT	23-MAR-1998	(first entry)
DE	Human oestrogen receptor PCR antisense primer E-7.	
XX	PCR primer: human; progesterone receptor; oestrogen; CD8; glucagon;	
KW	interleukin 2; insulin; parathyroid hormone; cholecystokinin;	
KW	enkephalin; adrenocorticotrophic hormone; thyroid stimulatory hormone;	
KW	detection; diagnosis; ss.	
XX	Synthetic.	
OS	Homo sapiens.	
XX	JP09187299-A.	
PN	22-JUL-1997.	
PD	05-JAN-1996;	96JP-0027222.
PF	05-JAN-1996;	96JP-0027222.
PR	05-JAN-1996;	96JP-0027222.
PA	(NIBI-) NIDPON BIOTHERAPY KK.	
XX	WPI; 1997-419410/39.	
DR	Polymerase chain reaction primers for detection of mRNA encoding	
XX	various human protein(s) - e.g. CD8, interleukin-2, parathyroid	
PT		

PT hormone, adrenocorticotrophic hormone and thyroid stimulating  
 PT hormone  
 XX  
 XX  
 PS Claim 1; Page 2; 4pp; Japanese.  
 XX  
 CC The present sequence represents a novel polymerase chain reaction  
 CC (PCR) primer. The PCR primers of the present specification are used  
 CC in the detection of mRNA encoding various human proteins. The mRNA  
 CC can be detected in a single cell.  
 XX  
 SQ Sequence 26 BP; 2 A; 8 C; 5 G; 11 T; 0 other;  
 Query Match 100.0%; Score 20; DB 10; Length 26;  
 Best Local Similarity 100.0%; Pred. No. 1.5;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CAAGCGCCAGAGAGATGATG 20  
 ||||||||||||||||  
 Db 20 CAAGCGCCAGAGAGATGATG 1  
 RESULT 2  
 AAX37808  
 ID AAX37808 standard; DNA; 521 BP.  
 XX  
 AC AAX37808;  
 XX  
 DF 09-JUL-1999 (first entry)  
 XX  
 DE PDM12 oestrogen receptor DNA binding domain fusion protein DNA.  
 XX  
 KW Peptide display carrier package; PDCP; chimeric protein; fusion protein;  
 KW screening; library; secreted peptide; oestrogen receptor;  
 KW DNA binding domain; ss.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 41..475  
 FT /tag= a  
 FT /product= "Oestrogen receptor DNA binding domain  
 fusion protein 5'-end"  
 XX  
 PN W09911785-A1.  
 XX  
 PD 11-MAR-1999.  
 XX  
 PF 02-SEP-1998; 98WO-GB02630.  
 XX  
 PR 02-SEP-1997; 97GB-0018455.  
 XX  
 PA (ROME-) ROWETT RES SERVICES LTD.  
 XX  
 PI McGregor D;  
 XX  
 DR WPI; 1999-302254/25.  
 DR P-PDB; AAY08223.  
 XX  
 PT New peptide display carrier package useful for screening nucleotide  
 PT libraries for sequences that encode peptides of interest  
 XX  
 PS Example 1; Fig 1; 103pp; English.  
 XX  
 CC This invention describes a novel peptide display carrier package (PDCP),  
 CC comprising a recombinant polynucleotide-chimeric protein complex, for  
 CC the isolation of peptide-encoding nucleic acid sequences from a DNA  
 CC expression library. The PDCP comprises a recombinant  
 CC polynucleotide-chimeric protein complex, where (a) the chimeric protein  
 CC has a nucleotide binding portion (NBP) and a target peptide portion  
 CC (TPP); (b) the recombinant polynucleotide comprises a specific sequence  
 CC motif which is specifically bound by the NBP; and (c) at least the  
 CC chimeric protein-encoding portion of the polynucleotide not bound by

CC the NBP is protected by a binding moiety. The PDCP is used in the  
 CC method for screening nucleotide libraries for sequences that encode  
 CC peptides of interest. Unlike prior art, the invention provides a method  
 CC to produce a peptide library where the DNA is protected from degradation,  
 CC and which allows production of secreted peptides with a free carboxy  
 CC terminus. This sequence encodes a fusion protein fragment used in the  
 CC method of the invention.  
 XX  
 SQ Sequence 521 BP; 153 A; 109 C; 155 G; 104 T; 0 other;  
 Query Match 100.0%; Score 20; DB 20; Length 521;  
 Best Local Similarity 100.0%; Pred. No. 2.4; Mismatches 0;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CAAGCGCCAGAGAGATGATG 20  
 ||||||||||||||||  
 Db 427 CAAGCGCCAGAGAGATGATG 446  
 RESULT 3  
 AAX37813  
 ID AAX37813 standard; DNA; 539 BP.  
 XX  
 AC AAX37813;  
 XX  
 DF 09-JUL-1999 (first entry)  
 XX  
 DE PDM16 oestrogen receptor DNA binding domain fusion protein DNA.  
 XX  
 KW Peptide display carrier package; PDCP; chimeric protein; fusion protein;  
 KW screening; library; secreted peptide; oestrogen receptor;  
 KW DNA binding domain; ss.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 41..481  
 FT /tag= a  
 FT /product= "Oestrogen receptor fusion protein"  
 XX  
 PN W09911785-A1.  
 XX  
 PD 11-MAR-1999.  
 XX  
 PF 02-SEP-1998; 98WO-GB02630.  
 XX  
 PR 02-SEP-1997; 97GB-0018455.  
 XX  
 PA (ROME-) ROWETT RES SERVICES LTD.  
 XX  
 PI McGregor D;  
 XX  
 DR WPI; 1999-302254/25.  
 DR P-PDB; AAY08228.  
 XX  
 PT New peptide display carrier package useful for screening nucleotide  
 PT libraries for sequences that encode peptides of interest  
 XX  
 PS Example 5; Fig 5; 103pp; English.  
 XX  
 CC This invention describes a novel peptide display carrier package (PDCP),  
 CC comprising a recombinant polynucleotide-chimeric protein complex, for  
 CC the isolation of peptide-encoding nucleic acid sequences from a DNA  
 CC expression library. The PDCP comprises a recombinant  
 CC polynucleotide-chimeric protein complex, where (a) the chimeric protein  
 CC has a nucleotide binding portion (NBP) and a target peptide portion  
 CC (TPP); (b) the recombinant polynucleotide comprises a specific sequence  
 CC motif which is specifically bound by the NBP; and (c) at least the  
 CC chimeric protein-encoding portion of the polynucleotide not bound by  
 CC the NBP is protected by a binding moiety. The PDCP is used in the  
 CC method for screening nucleotide libraries for sequences that encode  
 CC peptides of interest. Unlike prior art, the invention provides a method



CC to produce a peptide library where the DNA is protected from degradation,  
CC and which allows production of secreted peptides with a free carboxy  
CC terminus. This sequence encodes a fusion protein fragment used in the  
CC method of the invention.  
XX  
XX  
SQ Sequence 539 BP; 148 A; 118 C; 165 G; 108 T; 0 other;  
Query Match 100.0%; Score 20; DB 20; Length 539;  
Best Local Similarity 100.0%; Pred. No. 2.4;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CAAGCGCCAGAGAGATGATG 20  
DB 385 CAAGCGCCAGAGAGATGATG 404  
RESULT 4  
AA37812  
ID AA37812 standard; DNA; 566 BP.  
XX  
AC AA37812;  
XX  
DT 09-JUL-1999 (first entry)  
XX  
DE PDM14 oestrogen receptor DNA binding domain fusion protein DNA.  
XX  
KW Peptide display carrier package; PDCP; chimeric protein; fusion protein;  
KW screening; library; secreted peptide; oestrogen receptor;  
KW DNA binding domain; ss.  
XX  
OS Synthetic.  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 41..475  
FT /tag= a  
FT /product= "oestrogen receptor fusion protein"  
XX  
PN WO9911785-A1.  
XX  
PD 11-MAR-1999.  
XX  
PF 02-SEP-1998; 98WO-GB02630.  
XX  
PR 02-SEP-1997; 97GB-0018455.  
XX  
PA (ROME-) ROWETT RES SERVICES LTD.  
XX  
PI Mcgregor D;  
XX  
DR WPI; 1999-302254/25.  
DR P-PSDB; AAV08227.  
XX  
PT New peptide display carrier package useful for screening nucleotide  
PT libraries for sequences that encode peptides of interest  
XX  
PS Example 4; Fig 4; 103pp; English.  
XX  
CC This invention describes a novel peptide display carrier package (PDCP),  
CC comprising a recombinant polynucleotide-chimeric protein complex, for  
CC the isolation of peptide-encoding nucleic acid sequences from a DNA  
CC expression library. The PDCP comprises a recombinant  
CC polynucleotide-chimeric protein complex, where (a) the chimeric protein  
CC has a nucleotide binding portion (NBP) and a target peptide portion  
CC (TPP); (b) the recombinant polynucleotide comprises a specific sequence  
CC motif which is specifically bound by the NBP; and (c) at least the  
CC chimeric protein-encoding portion of the polynucleotide not bound by  
CC the NBP is protected by a binding moiety. The PDCP is used in the  
CC method for screening nucleotide libraries for sequences that encode  
CC peptides of interest. Unlike prior art, the invention provides a method  
CC to produce a peptide library where the DNA is protected from degradation,  
CC and which allows production of secreted peptides with a free carboxy  
CC terminus. This sequence encodes a fusion protein fragment used in the

CC method of the invention.  
XX  
XX  
SQ Sequence 566 BP; 164 A; 118 C; 169 G; 115 T; 0 other;  
Query Match 100.0%; Score 20; DB 20; Length 566;  
Best Local Similarity 100.0%; Pred. No. 2.4;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CAAGCGCCAGAGAGATGATG 20  
DB 427 CAAGCGCCAGAGAGATGATG 446  
RESULT 5  
AA29924  
ID AA29924 standard; DNA; 1223 BP.  
XX  
AC AA29924;  
XX  
DT 04-APR-2001 (first entry)  
XX  
DE Human estrogen receptor alpha isoform #1 DNA.  
XX  
KW Human; estrogen receptor alpha; cancer; osteoporosis; bone;  
KW Alzheimer's; cardiovascular; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200100823-A1.  
XX  
PD 04-JAN-2001.  
XX  
PF 27-JUN-2000; 2000WO-EP05981.  
XX  
PR 29-JUN-1999; 99IT-WI01433.  
XX  
PA (EURO-) EURO MOLECULAR BIOLOGY LAB.  
XX  
PI Gannon F, Dengler S, Flouriot G;  
XX  
DR WPI; 2001-137955/14.  
XX  
XX  
PT Novel isoforms of human estrogen receptor alpha useful for preparing  
PT therapeutic agents for treating cancer, osteoporosis, Alzheimer's  
PT disease and cardiovascular diseases  
XX  
PS Claim 1; Pag e44; 53pp; English.  
XX  
CC The present invention relates to a human estrogen receptor  
CC (HER)-alpha isoform. Molecules which modulate the activity of  
CC the estrogen receptor are useful for the preparation of therapeutic  
CC agents for treating cancer, osteoporosis and other bone disorders,  
CC Alzheimer's disease and cardiovascular diseases.  
XX  
SQ Sequence 1223 BP; 303 A; 316 C; 341 G; 263 T; 0 other;  
Query Match 100.0%; Score 20; DB 22; Length 1223;  
Best Local Similarity 100.0%; Pred. No. 2.7;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CAAGCGCCAGAGAGATGATG 20  
DB 236 CAAGCGCCAGAGAGATGATG 255  
RESULT 6  
AA29925  
ID AA29925 standard; DNA; 1374 BP.  
XX  
AC AA29925;  
XX  
DT 04-APR-2001 (first entry)  
XX

DE Human estrogen receptor alpha isoform #2 DNA.  
 KW Human; estrogen receptor alpha; cancer; osteoporosis; bone;  
 KW Alzheimer's; cardiovascular; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200100823-A1.  
 XX  
 PD 04-JAN-2001.  
 XX  
 PF 27-JUN-2000; 2000WO-EP05981.  
 XX  
 PR 29-JUN-1999; 99IT-MI01433.  
 XX  
 PA (EUMO-) EURO MOLECULAR BIOLOGY LAB.  
 XX  
 PI Gannon F, Dengler S, Flouriot G;  
 XX  
 DR WPI; 2001-137955/14.  
 XX  
 PT Novel isoforms of human estrogen receptor alpha useful for preparing  
 PT therapeutic agents for treating cancer, osteoporosis, Alzheimer's  
 PT disease and cardiovascular diseases -  
 XX  
 PS Claim 1; Page 44-45; 53pp; English.  
 XX  
 CC The present invention relates to a human estrogen receptor  
 CC (HER) alpha isoform. Molecules which modulate the activity of  
 CC the estrogen receptor are useful for the preparation of therapeutic  
 CC agents for treating cancer, osteoporosis and other bone disorders,  
 CC Alzheimer's disease and cardiovascular diseases.  
 XX  
 SO Sequence 1374 BP; 349 A; 346 C; 386 G; 293 T; 0 other;  
 OY  
 1 CAAGGCCAGAGATGATG 20  
 ||||||||||||||||  
 DB 387 CAAGGCCAGAGATGATG 406  
 OY  
 1 CAAGGCCAGAGATGATG 20  
 ||||||||||||||||  
 DB 387 CAAGGCCAGAGATGATG 406  
 RESULT 7  
 AAC86919  
 ID AAC86919 standard; DNA; 1380 BP.  
 XX  
 AC AAC86919;  
 XX  
 DT 02-APR-2001 (first entry)  
 XX  
 DE Nucleotide sequence of the DNA binding domain of estrogen receptor.  
 XX  
 KW Ribozyme; oestrogen-dependent tumour; cell proliferation; glucocorticoid;  
 KW DNA-binding domain; oestrogen receptor; cancer treatment; breast cancer;  
 KW ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200074485-A1.  
 XX  
 PD 14-DEC-2000.  
 XX  
 PF 02-JUN-2000; 2000WO-US15243.  
 XX  
 PR 04-JUN-1999; 99US-0137470.  
 XX  
 PA (TEXA) UNIV TEXAS.  
 XX  
 PI Roy AK, Lavyrovsky Y, Tyagi RK, Song CS, Chatterjee B;  
 XX  
 DR WPI; 2001-061633/07.

XX  
 PT Ribozyme having a high substrate specificity for an mRNA encoding a  
 PT DNA-binding domain of human estrogen receptor, useful for inhibiting  
 PT estrogen-dependent tumour cell proliferation, particularly breast cancer  
 PT  
 XX  
 PS Claim 1; Page 6-7; 49pp; English.  
 XX  
 CC The specification describes a ribozyme capable of inhibiting  
 CC oestrogen-dependent tumour cell proliferation and having a high  
 CC substrate specificity for an mRNA sequence encoding a DNA-binding  
 CC domain of human estrogen receptor. The ribozyme is free of endonuclease  
 CC activity for an mRNA having a DNA binding domain of a glucocorticoid.  
 CC The estrogen receptor site-specific ribozymes are useful for cancer  
 CC treatment and therapies, especially for inhibiting estrogen-dependent  
 CC tumour cell proliferation, particularly breast cancer. The present  
 CC sequence represents the DNA binding domain of a human estrogen receptor.  
 XX  
 SO Sequence 1380 BP; 320 A; 385 C; 400 G; 275 T; 0 other;  
 OY  
 1 CAAGGCCAGAGATGATG 20  
 ||||||||||||||||  
 DB 733 CAAGGCCAGAGATGATG 752  
 OY  
 1 CAAGGCCAGAGATGATG 20  
 ||||||||||||||||  
 DB 733 CAAGGCCAGAGATGATG 752  
 RESULT 8  
 ABL57497  
 ID ABL57497 standard; cDNA; 1788 BP.  
 XX  
 AC ABL57497;  
 XX  
 DT 22-AUG-2002 (first entry)  
 XX  
 DE Human nuclear oestrogen receptor alpha coding sequence.  
 XX  
 KW Oestrogen; receptor; human; transgenic mouse; cytostatic;  
 KW antiinflammatory; antidiabetic; endocrine; anorectic; hepatotropic;  
 KW gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key location/Qualifiers  
 FH 1..1788  
 FT CDS /tag= a  
 FT /product= "Nuclear oestrogen receptor"  
 FT  
 PN WO200228175-A2.  
 XX  
 PD 11-APR-2002.  
 XX  
 PF 28-SEP-2001; 2001WO-IB02246.  
 XX  
 PR 03-OCT-2000; 2000FR-0012570.  
 XX  
 PR 11-MAY-2001; 2001US-0853033.  
 XX  
 PA (ADER-) ADEREGEM ASSOC DEV RECH EN GENETIQ.  
 XX  
 PI Chandon P, Metzger D;  
 XX  
 DR WPI; 2002-463217/49.  
 DR P-PSDB; ABB76378.  
 XX  
 PT A transgenic mouse, useful in screening for medicaments for the  
 PT treatment of e.g., diabetes or skin cancers, comprises a fusion protein  
 PT between a recombinase Cre, and a modified ligand binding domain of the  
 PT nuclear estrogen receptor alpha -  
 XX  
 PS Disclosure; Page 128-130; 149pp; English.

CC The present sequence is the coding sequence of the human nuclear  
CC oestrogen receptor alpha (1). The invention relates to a non-human  
CC metazoan organism, especially a transgenic mouse, characterised in  
CC that at least one cell comprises: (i) a fusion protein formed from  
CC a recombinase (Cre), a hinge region (preferably derived from human  
CC (1)), and a modified ligand binding domain (LBD) of a nuclear  
CC oestrogen receptor, such as (1), its fragment or variant; and (ii)  
CC one or more genes or DNA sequences of interest belonging to the  
CC genome of the organism, into which one or more recognition sites  
CC of the recombinase protein are inserted. The (1) LBD domain is  
CC preferably modified by a G521R, G400V, or M543A/T544A mutation.  
CC The fusion protein has negligible, or even zero, recombinase  
CC activity in the presence of a natural ligand such as oestradiol;  
CC but recombinase activity is induced by a small quantity of a  
CC synthetic ligand that has antioestrogenic activity, e.g. tamoxifen  
CC or 4-hydroxytamoxifen. The metazoan organism or its cells, such  
CC as epidermal cells, hepatocytes or adipocytes, are useful in  
CC carrying out a spatiotemporally controlled site-specific  
CC recombination of a DNA sequence of interest in its natural  
CC chromatin environment. It is also used in screening of medicaments  
CC for pathological conditions associated with an alteration of the  
CC expression and/or function of the DNA sequence of interest, such as  
CC skin cancer, inflammation, diabetes, alopecia, obesity, or in  
CC promoting hepatic regeneration.

SO Sequence 1788 BP; 418 A; 515 C; 508 G; 347 T; 0 other;

Query Match 100.0%; Score 20; DB 24; Length 1788;  
Best Local Similarity 100.0%; Pred. No. 2.9;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGGCCGAGAGATGATG 20  
DB 801 CAAGGCCGAGAGATGATG 820

RESULT 9  
ID AANT0880 standard; cDNA; 2092 BP.

XX AANT0880;  
XX 24-FEB-1991 (first entry).  
XX cDNA encoding human oestrogen receptor protein.  
XX DE  
XX KM Human oestrogen receptor protein; ss cDNA; expression system; assay;  
XX KW steroid receptor protein;  
XX OS Homo sapiens.  
XX XX  
XX FH Key Location/Qualifiers  
XX FT CDS 293..1280  
XX FT /\*tag= a  
XX PN W08705049-A.  
XX PD 27-AUG-1987.  
XX PF 18-FEB-1987; 87WO-US00341.  
XX PR 20-FEB-1986; 86US-0833829.  
XX PA (CALB-) CALIFORNIA BIOTECHNOLOGY INC.  
XX PI Shine J;  
XX WPI; 1987-250213/35.  
XX P-PSDB; AAF70543.  
XX Expression system for vertebrate steroid receptor protein - comprising  
PT DNA sequence encoding the protein linked to control sequences in  
PT eucaryotic hosts.

XX Disclosure; Fig. 1-1 - 1-2; 26pp; English.  
XX The cDNA can be used in an expression system to express human  
XX oestrogen receptor protein. The coding sequence is operably linked to  
XX control sequences compatible with eukaryotic host cells. This method  
XX allows expression under conditions which favour appropriate post-  
XX transcriptional processing. It produces large amts. of purified protein  
XX useful in the design of agonist and antagonist cpts., for the study of  
XX the mechanism of action of the steroid binding proteins in general,  
XX and for use in diagnostic assays for the proteins or antibodies to  
XX them. These assays are important in, eg the diagnosis of tumour  
XX sensitivities to steroid metabolism. Suitable host cells are VERO,  
XX HeLa and CHO cells.

SO Sequence 2092 BP; 473 A; 607 C; 591 G; 421 T; 0 other;

Query Match 100.0%; Score 20; DB 8; Length 2092;  
Best Local Similarity 100.0%; Pred. No. 3;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGGCCGAGAGATGATG 20  
DB 1093 CAAGGCCGAGAGATGATG 1112

RESULT 10  
ID AAC86920 standard; cDNA; 2092 BP.

XX AAC86920;  
XX 02-APR-2001 (first entry)  
XX DE  
XX KM Nucleotide sequence of the human oestrogen receptor cDNA.  
XX KM Ribozyme; oestrogen-dependent tumour; cell proliferation; glucocorticoid;  
XX KM DNA-binding domain; oestrogen receptor; cancer treatment; breast cancer;  
XX KM ss.  
XX OS Homo sapiens.  
XX PN W0200074485-A1.  
XX PD 14-DEC-2000.  
XX PF 02-JUN-2000; 2000WO-US15243.  
XX PR 04-JUN-1999; 99US-0137470.  
XX PA (TEXA ) UNIV TEXAS.  
XX PI Roy AK, Lavyrovsky Y, Tyagi RK, Song CS, Chatterjee B;  
XX WPI; 2001-061633/07.  
XX PT Ribozyme having a high substrate specificity for an mRNA encoding a  
PT DNA-binding domain of human estrogen receptor, useful for inhibiting  
PT estrogen-dependent tumour cell proliferation, particularly breast cancer  
PT -  
PS Disclosure; Page 8-9; 49pp; English.  
XX The specification describes a ribozyme capable of inhibiting  
XX oestrogen-dependent tumour cell proliferation and having a high  
XX substrate specificity for an mRNA sequence encoding a DNA-binding  
XX domain of human oestrogen receptor. The ribozyme is free of endonuclease  
XX activity for an mRNA having a DNA binding domain of a glucocorticoid.  
XX The oestrogen receptor site-specific ribozymes are useful for cancer  
XX treatment and therapies, especially for inhibiting oestrogen-dependent  
XX tumour cell proliferation, particularly breast cancer. The present  
XX sequence represents the human oestrogen receptor cDNA.

CC detection of target molecule. (I) is also useful in diagnostic applications and drug optimization. The present researches need the

XX

```

XX DE SSR-LBD fusion polypeptide encoding DNA.
XX XX
XX KW Fusion protein; recombinase; nuclear receptor; ligand binding domain;
XX KW site-specific recombination; SSR-LBD fusion protein; FLP recombinase;
XX KW estrogen binding domain; ss.
XX OS Synthetic.
XX FH Key Location/Qualifiers
XX FT CDS 1..2322
XX FT /tag= a
XX FT /product= "SSR-LBD fusion protein"
XX FT misc_feature 1..1269
XX FT /tag= b
XX FT /note= "sequence coding for FLP recombinase domain"
XX FT misc_feature 1270..1284
XX FT /tag= c
XX FT /note= "sequence coding for the linker peptide"
XX FT misc_feature 1285..2322
XX FT /tag= d
XX FT /note= "sequence coding for estrogen binding domain"
XX PN WO9500555-A1.
XX PD 05-JAN-1995.
XX PF 28-JUN-1994; 94MO-EP02088.
XX PR 28-JUN-1993; 93EP-0110298.
XX PA (EMBL-) EMBL EURO MOLECULAR BIOLOGY LAB.
XX PI Stewart F;
XX WP1: 1999-337364/28.
XX P-PSDB; AAY21544.
XX PT New fusion protein - comprising a recombinase protein and nuclear
XX PT receptor, useful for regulation of recombination
XX PS Disclosure; Fig 4; 28pp; English.
XX CC The invention describes a fusion protein that comprises a recombinase
XX CC protein or a component of a recombinase complex fused to part of or all
XX CC of a nuclear receptor, where recombinase activity is inhibited in the
XX CC absence of amino acids binding the ligand binding domain of the receptor,
XX CC and activity is altered or induced by ligand binding. The new fusion
XX CC protein is useful for regulating site-specific recombination in cells or
XX CC organisms, for determining the effect of mutations in the ligand binding
XX CC domain of a nuclear receptor on its ability to bind a ligand, and for
XX CC assessing binding of a ligand to the ligand binding domain. Prior art
XX CC methods of recombinase regulation rely on inducible heat-shock promoters,
XX CC but these are only applicable in flies, and no suitable counterpart has
XX CC been found. The present sequence represents a nucleotide sequence coding
XX CC for the SSR-LBD fusion polypeptide. The fusion protein consists of an FLP
XX CC recombinase domain linked by a linker peptide to an estrogen binding
XX CC domain.
XX SQ Sequence 2322 BP; 703 A; 505 C; 545 G; 569 T; 0 other;

Query Match 100.0%; Score 20; DB 20; Length 2322;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAAGGCCGAGAGATGATG 20
DB 1335 CAAGGCCGAGAGATGATG 1354

```

RESULT 14  
AA06873  
ID AA06873 standard; DNA; 4963 BP.

```

XX AC AA06873;
XX DT 13-MAR-1996 (first entry)
XX DE PhCMV*-1 promoter and human oestrogen receptor.
XX KW Transactivator; tTA; tet operator; tetO; gene expression;
XX KW tetracycline-responsive promoter; PhCMV*-1; human cytomegalovirus;
XX KW transcription activation; eukaryotic cell; estrogen receptor;
XX KW ds; ss.
XX OS Chimeric synthetic;
XX OS Chimeric Homo sapiens.
XX PN US5464758-A.
XX PD 07-NOV-1995.
XX PF 14-JUN-1993; 93US-0076726.
XX PR 14-JUN-1993; 93US-0076726.
XX PA (BUJA/) BUJARD H.
XX PA (GOSS/) GOSSSEN M.
XX PI Bujard H, Gossen M;
XX WP1: 1995-392612/50.
XX PT Polynucleotide encoding transactivator fusion protein contg. tet
XX PT repressor - used to control expression of gene regulated by minimal
XX PT promoter linked to tet operon, and vectors and cells where gene
XX PT expression is regulated by tetracycline
XX PS Disclosure; Fig 10A-B; 37pp; English.
XX CC cDNA coding for the human oestrogen receptor was placed under
XX CC control of the PhCMV*-1 promoter (see AA06869), resulting in
XX CC tetracycline-regulated expression vector pUHPgr-3 contg. the
XX CC sequence given in AA06873. The construct was transfected into
XX CC eukaryotic cells expressing the tTA transactivator (AA06867) to
XX CC allow tetracycline-regulated expression of the receptor.
XX SQ Sequence 4963 BP; 1259 A; 1300 C; 1278 G; 1126 T; 0 other;

Query Match 100.0%; Score 20; DB 16; Length 4963;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAAGGCCGAGAGATGATG 20
DB 1266 CAAGGCCGAGAGATGATG 1285

```

RESULT 15  
AA076270  
ID AA076270 standard; DNA; 4963 BP.  
AC AA076270;  
DT 17-JUL-1995 (first entry)  
DE PhCMV\*-1-controlled progesterone receptor.  
KW tTA; transactivator; tetracycline-controlled transactivator;  
KW conditional inactivation; homologous recombination; gene expression;  
KW gene regulation; gene therapy; tetracycline-resistance; tetR;  
KW transgenic animal; PhCMV\*-1; promoter; tetO; CMV;  
KW progesterone receptor; ds.  
XX OS Human cytomegalovirus Towne.

PN W09429442-A.  
 XX  
 PD 22-DEC-1994.  
 XX  
 PF 14-JUN-1994; 94MO-US06734.  
 XX  
 PR 14-JUN-1993; 93US-0076327.  
 XX  
 PA (BADI ) BASF AG.  
 XX  
 PI Bujard H, Gossen M, Salfeld JG, Voss JW;  
 XX  
 DR WPI; 1995-036472/05.  
 XX  
 PT Regulatory systems using tetracycline-controllable transactivator  
 PT (ttr) - useful for conditional inactivation or modulation of  
 PT gene expression in a host cell or animal  
 XX  
 PS Disclosure; Page 54-58; 103pp; English.  
 XX  
 CC The minimal promoter PhCMV\*-1 is based on the human CMV  
 CC intermediate-early promoter (with enhancer region removed), fused  
 CC to a tetracycline operator, tetrO. The promoter is used to  
 CC express tTA transactivator in host cells. In clone pUHD BGR4  
 CC (given in AAQ76269), cDNA for rabbit progesterone receptor is under  
 CC the control of PhCMV\*-1.  
 XX  
 SO Sequence 4963 BP; 1259 A; 1298 C; 1280 G; 1126 T; 0 other;  
 Query Match 100.0%; Score 20; DB 16; Length 4963;  
 Best Local Similarity 100.0%; Pred. No. 3.4;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CAAGGCGCAGAGAGATGATG 20  
 Db 1266 CAAGGCGCAGAGAGATGATG 1285

Search completed: May 16, 2003, 15:58:25  
 Job time : 116.126 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Comphen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 16, 2003, 19:53:58 ; Search time 23.1788 seconds

(without alignments)  
264.618 Million cell updates/sec

Title: US-10-052-092-15

Sequence: 1 caagcgcagagagatgatg 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Length	ID	Description
1	20	100.0	2222 3	US-08-564-264-2
2	20	100.0	4963 1	US-08-076-726-16
3	20	100.0	4963 2	US-08-260-452-9
4	20	100.0	4963 2	US-08-481-970-9
5	20	100.0	4963 2	US-08-897-719-9
6	20	100.0	4963 4	US-09-163-269-9
7	20	100.0	6450 4	US-09-041-886-34
8	20	100.0	6450 4	US-08-453-998-1
9	15.8	79.0	2328 1	US-08-688-649-38
10	15.8	79.0	2328 4	US-09-715-524B-1
11	15.8	79.0	3499 1	US-07-966-278-2
12	15.8	79.0	3499 1	US-08-424-921-2
13	15.8	79.0	3499 2	US-08-556-355A-2
14	15.8	79.0	3499 2	US-07-803-627A-2
15	15.4	77.0	1542 3	US-09-089-879-9
16	15.4	77.0	3640 4	US-09-221-017B-1047
17	15.4	77.0	35060 3	US-08-814-095-7
18	15.2	76.0	737 2	US-08-861-549-2
19	15.2	76.0	1308 1	US-08-518-878B-57
20	15.2	76.0	1308 2	US-08-807-861A-57
21	15.2	76.0	1308 3	US-09-210-681-57
22	15.2	76.0	1308 3	US-08-946-719A-57
23	15.2	76.0	3870 1	US-08-138-641-1
24	15.2	76.0	3870 1	US-08-138-133-1
25	15.2	76.0	3893 1	US-08-138-641-3
26	15.2	76.0	3893 1	US-08-138-133-3
27	15.2	76.0	4198 4	US-09-586-719-11

28	14.8	74.0	575 4	US-09-385-982-450	Sequence 450, App
29	14.8	74.0	1776 4	US-09-149-476-59	Sequence 59, Appl
30	14.8	74.0	1791 4	US-09-149-476-226	Sequence 226, App
31	14.8	74.0	5455 1	US-08-342-930-1	Sequence 1, Appl1
32	14.8	74.0	9848 4	US-09-385-222A-3	Sequence 3, Appl1
33	14.8	74.0	10163 2	US-08-659-251-1	Sequence 1, Appl1
34	14.8	74.0	10163 4	US-09-256-490-1	Sequence 1, Appl1
35	14.8	74.0	10163 5	PCR-US96-11445-1	Sequence 1, Appl1
36	14.4	72.0	810 4	US-09-230-041-1	Sequence 1, Appl1
37	14.2	71.0	30 3	US-09-167-717-6	Sequence 6, Appl1
38	14.2	71.0	750 3	US-09-167-717-4	Sequence 4, Appl1
39	14.2	71.0	864 4	US-09-439-261-12	Sequence 12, Appl1
40	14.2	71.0	864 4	US-09-227-613-13	Sequence 13, Appl1
41	14.2	71.0	918 4	US-09-439-261-5	Sequence 5, Appl1
42	14.2	71.0	918 4	US-09-227-613-5	Sequence 5, Appl1
43	14.2	71.0	960 4	US-09-439-261-36	Sequence 36, Appl1
44	14.2	71.0	960 4	US-09-227-613-35	Sequence 35, Appl1
45	14.2	71.0	990 4	US-09-439-261-35	Sequence 35, Appl1

## ALIGNMENTS

RESULT 1  
US-08-564-264-2  
; Sequence 2, Application US/08564264  
; Patent No. 6040430  
; GENERAL INFORMATION:  
; APPLICANT: STEWART, FRANCIS  
; TITLE OF INVENTION: REGULATION OF SITE-SPECIFIC  
; TITLE OF INVENTION: RECOMBINATION BY SITE-SPECIFIC RECOMBINASE/NUCLEAR  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Mikaido, Marmelstein, Murray & Oram  
; STREET: 655 Fifteenth Street N.W. Suite 330  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-5701  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/564,264  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP94/02088  
; FILING DATE: 28-JUN-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 93 110 298.2  
; FILING DATE: 28-JUN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murray, Robert B.  
; REGISTRATION NUMBER: 22,980  
; REFERENCE/DOCKET NUMBER: P564-5019  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)638-5000  
; TELEFAX: (202)638-4810  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2322 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1-1269)  
; OTHER INFORMATION: /note="FLP recombinase domain."

```
NAME/KEY: misc.feature
LOCATION: (1270-1284)
OTHER INFORMATION: /note= "Linker peptide."
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1285-2322)
OTHER INFORMATION: /note= "Estrogen binding domain."
US-08-564-264-2

Query Match
Best Local Similarity 100.0%; Score 20; DB 3; Length 2322;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGCGCCAGAGAGATGATG 20
DB 1335 CAAGCGCCAGAGAGATGATG 1354

RESULT 2
US-08-076-726-16
Sequence 16, Application US/08076726
Patent No. 5464758
GENERAL INFORMATION:
APPLICANT: Gossen, Manfred
APPLICANT: Bujard, Hermann
TITLE OF INVENTION: Tight Control of Gene Expression in
Eucaryotic Cells by Tetracycline-responsive Promoters
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, N.W.
CITY: Washington
STATE: District of Columbia
COUNTRY: United States of America
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/076,726
FILING DATE: 14-JUN-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 0942,2490001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)371-2600
TELEFAX: (202)371-2540
TELEX: 248636 SSK
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 4963 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
US-08-076-726-16

Query Match
Best Local Similarity 100.0%; Score 20; DB 1; Length 4963;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGCGCCAGAGAGATGATG 20
DB 1266 CAAGCGCCAGAGAGATGATG 1285

RESULT 3
US-08-260-452-9
Sequence 9, Application US/08260452
Patent No. 5650298
```

```
GENERAL INFORMATION:
APPLICANT: Gossen, Manfred
APPLICANT: Bujard, Hermann
APPLICANT: Salfeld, Jochem
APPLICANT: Voss, Jeffrey
TITLE OF INVENTION: Tight Control of Gene Expression in Eucaryotic
Cells by Tetracycline-responsive Promoters
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lanhive & Cockfield
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/260,452
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/076,327
ATTORNEY/AGENT INFORMATION:
NAME: Giulio A. DeConti, Jr.
REGISTRATION NUMBER: 31,503
REFERENCE/DOCKET NUMBER: BBI-013CP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 4963 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Human cytomegalovirus
IMMEDIATE SOURCE:
CLONE: pTHD BGR4
US-08-260-452-9

Query Match
Best Local Similarity 100.0%; Score 20; DB 1; Length 4963;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGCGCCAGAGAGATGATG 20
DB 1266 CAAGCGCCAGAGAGATGATG 1285

RESULT 4
US-08-481-970-9
Sequence 9, Application US/08481970
Patent No. 5859310
GENERAL INFORMATION:
APPLICANT: Gossen, Manfred
APPLICANT: Bujard, Hermann
APPLICANT: Salfeld, Jochem
APPLICANT: Voss, Jeffrey
TITLE OF INVENTION: Animal Transgenic for a Tetracycline-Controlled
Transcri
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lanhive & Cockfield
STREET: 60 State Street, Suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
```



```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/481.970
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/260.452
FILING DATE: 14-JUN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/076.327
FILING DATE: 14-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Giulio A. DeConti, Jr.
REGISTRATION NUMBER: 31,503
REFERENCE/DOCKET NUMBER: BBI-013CP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 4963 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Human cytomegalovirus
IMMEDIATE SOURCE:
CLONE: pUHD BGR4
US-08-481-970-9

Query Match
Best Local Similarity 100.0%; Score 20; DB 2; Length 4963;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGCGCCAGAGAGATGATG 20
Db 1266 CAAGCGCCAGAGAGATGATG 1285

RESULT 5
US-08-897-719-9
Sequence 9, Application US/08897719
Patent No. 5922927
GENERAL INFORMATION:
APPLICANT: Gossen, Manfred
APPLICANT: Bajard, Hermann
APPLICANT: Salfeld, Jochem
APPLICANT: Voss, Jeffrey
TITLE OF INVENTION: Tight Control of Gene Expression in Eucaryotic
TITLE OF INVENTION: Cells by Tetracycline-responsive Promoters
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lahive & Cockfield
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/897.719
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
```

```
APPLICATION NUMBER: 08/260.452
FILING DATE:
APPLICATION NUMBER: 08/076.327
FILING DATE: 14-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Giulio A. DeConti, Jr.
REGISTRATION NUMBER: 31,503
REFERENCE/DOCKET NUMBER: BBI-013CP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 4963 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Human cytomegalovirus
IMMEDIATE SOURCE:
CLONE: pUHD BGR4
US-08-897-719-9

Query Match
Best Local Similarity 100.0%; Score 20; DB 2; Length 4963;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGCGCCAGAGAGATGATG 20
Db 1266 CAAGCGCCAGAGAGATGATG 1285

RESULT 6
US-09-163-269-9
Sequence 9, Application US/09163269
Patent No. 6252136
GENERAL INFORMATION:
APPLICANT: Gossen, Manfred
APPLICANT: Bajard, Hermann
APPLICANT: Salfeld, Jochem
APPLICANT: Voss, Jeffrey
TITLE OF INVENTION: Animal Transgenic for a Tetracycline-Controlled
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lahive & Cockfield
STREET: 60 State Street, Suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/163.269
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/481.970
FILING DATE:
APPLICATION NUMBER: 08/076.327
FILING DATE: 14-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Giulio A. DeConti, Jr.
REGISTRATION NUMBER: 31,503
REFERENCE/DOCKET NUMBER: BBI-013CP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
Transcr
```

INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4963 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Human cytomegalovirus  
IMMEDIATE SOURCE:  
CLONE: PCHD B6R4  
US-09-163-269-9

Query Match 100.0%; Score 20; DB 4; Length 4963;  
Best Local Similarity 100.0%; Pred. No. 0.58;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CAAGCCGACAGAGATGATG 20  
1266 CAAGCCGACAGAGATGATG 1285

RESULT 7  
US-09-041-886-34  
Sequence 34, Application US/09041886  
Patent No. 6235872  
GENERAL INFORMATION:  
APPLICANT: Bredesen, Dale E.  
TITLE OF INVENTION: Proapoptotic Peptides, Dependence  
TITLE OF INVENTION: Polypeptides and Methods of Use  
NUMBER OF SEQUENCES: 72  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Campbell & Flores LLP  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: United States  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/041,886  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-LT 2626  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6450 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 361..2146  
US-09-041-886-34  
Query Match 100.0%; Score 20; DB 4; Length 6450;  
Best Local Similarity 100.0%; Pred. No. 0.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 CAAGCCGACAGAGATGATG 20  
|||||

Db 1161 CAAGCCGACAGAGATGATG 1180

RESULT 8  
US-08-453-998-1  
Sequence 1, Application US/08453998  
Patent No. 6444438  
GENERAL INFORMATION:  
APPLICANT: CHABON, PIERRE  
APPLICANT: METZGER, DANIEL  
APPLICANT: WHITE, JOHN  
TITLE OF INVENTION: METHOD FOR THE PREPARATION OF A PROTEIN  
TITLE OF INVENTION: BY YEASTS USING AN INDUCIBLE SYSTEM, VECTORS AND  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: CUSHMAN, DAREY & CUSHMAN  
STREET: 1100 NEW YORK AVE., N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/453,998  
FILING DATE:  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/161,064  
FILING DATE: 03-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: CHAPIN, MARIANA K.  
REGISTRATION NUMBER: 35,843  
REFERENCE/DOCKET NUMBER: 1037/98493  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-861-3711  
TELEFAX: 202-822-0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6450 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-453-998-1

Query Match 100.0%; Score 20; DB 4; Length 6450;  
Best Local Similarity 100.0%; Pred. No. 0.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CAAGCCGACAGAGATGATG 20  
1161 CAAGCCGACAGAGATGATG 1180

RESULT 9  
US-08-688-649-38  
Sequence 38, Application US/08688649  
Patent No. 5827716  
GENERAL INFORMATION:  
APPLICANT: MARONE, JOSEPH A.  
TITLE OF INVENTION: MODIFIED POL-II TYPE DNA  
TITLE OF INVENTION: POLYMERASES  
NUMBER OF SEQUENCES: 38  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: LYON & LYON  
STREET: 633 West Fifth Street  
STREET: Suite 4700

```

; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/688, 649
; FILING DATE: Filed Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 220/281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2328 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-688-649-38

Query Match          79.0%; Score 15.8; DB 1; Length 2328;
Best Local Similarity 89.5%; Pred. No. 57;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 CAAGGCCAGAGAGATGAT 19
DB 554 CAAGCGAGAGAGATGAT 572

RESULT 10
US-09-715-524B-1
; Sequence 1, Application US/09715524B
; Patent No. 633183
; GENERAL INFORMATION:
; APPLICANT: Evans, Steven
; APPLICANT: Mamone, Joseph
; APPLICANT: Davis, Maria
; APPLICANT: Connolly, Bernard
; TITLE OF INVENTION: Improving Dideoxynucleotide-Triphosphate Utilization by
; TITLE OF INVENTION: the Hyper-Thermophilic DNA Polymerase from the Archaeon
; FILE REFERENCE: PB-9965
; CURRENT APPLICATION NUMBER: US/09/715,524B
; CURRENT FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/167,066
; PRIOR FILING DATE: 1999-11-23
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentia Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2328
; TYPE: DNA
; ORGANISM: Pyrococcus furiosus
;
US-09-715-524B-1

Query Match          79.0%; Score 15.8; DB 4; Length 2328;
Best Local Similarity 89.5%; Pred. No. 57;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 CAAGGCCAGAGAGATGAT 19
DB 554 CAAGCGAGAGAGATGAT 572

```

```

DB 554 CAAGCGAGAGAGATGAT 572

RESULT 11
US-07-966-278-2
; Sequence 2, Application US/07966278
; Patent No. 5489523
; GENERAL INFORMATION:
; APPLICANT: Mathur, Eric A
; TITLE OF INVENTION: EXONUCLEASE-DEFICIENT THERMOSTABLE
; TITLE OF INVENTION: PYROCOCUS FURIOSUS DNA POLYMERASE I
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Thomas Fitting
; STREET: 12526 High Bluff Road, Suite 300
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92130
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentia Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/966,278
; FILING DATE: 19921226
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitting, Thomas
; REGISTRATION NUMBER: 34,163
; REFERENCE/DOCKET NUMBER: STG0133P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-792-3680
; TELEFAX: 619-792-8677
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3499 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..223
; NAME/KEY: CDS
; LOCATION: 224..2551
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: 2552..3499
;
US-07-966-278-2

Query Match          79.0%; Score 15.8; DB 1; Length 3499;
Best Local Similarity 89.5%; Pred. No. 60;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 CAAGGCCAGAGAGATGAT 19
DB 777 CAAGCGAGAGAGATGAT 795

RESULT 12
US-08-424-921-2
; Sequence 2, Application US/08424921
; Patent No. 5545552
; GENERAL INFORMATION:
; APPLICANT: Mathur, Eric A

```

;; TITLE OF INVENTION: PURIFIED THERMOSTABLE PYROCOCCLUS  
;; TITLE OF INVENTION: FURIOSUS DNA POLYMERASE I  
;; NUMBER OF SEQUENCES: 6  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Bingham & Filling  
;; STREET: 12526 High Bluff Road, Suite 300  
;; CITY: San Diego  
;; STATE: CA  
;; COUNTRY: USA  
;; ZIP: 92130  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentln Release #1.0, Version #1.25  
;;  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/424,921  
;; FILING DATE: 19-APR-1995  
;; CLASSIFICATION: 435  
;;  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/803,627  
;; FILING DATE: 02-DEC-1991  
;; APPLICATION NUMBER: US 07/620,568  
;; FILING DATE: 03-DEC-1990  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/657,073  
;; FILING DATE: 19-FEB-1991  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/776,552  
;; FILING DATE: 15-OCT-1991  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Filling, Thomas  
;; REGISTRATION NUMBER: 34,163  
;; REFERENCE/DOCKET NUMBER: STG0100P  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 619-792-3680  
;; TELEFAX: 619-792-8477  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 3499 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; HYPOTHEetical: NO  
;; ANTI-SENSE: NO  
;; FEATURE:  
;; NAME/KEY: 5'UTR  
;; LOCATION: 1..223  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 224..2551  
;; FEATURE:  
;; NAME/KEY: 3'UTR  
;; LOCATION: 2552..3499  
;;  
;; US-08-424-921-2  
;;  
Query Match 79.0%; Score 15.8; DB 1; Length 3499;  
Best Local Similarity 89.5%; Pred. No. 60;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

;; TITLE OF INVENTION: furiosus DNA Polymerase I  
;; NUMBER OF SEQUENCES: 6  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Finegan, Henderson, Farabow, Garrett &  
;; STREET: 1300 I Street, N.W.  
;; CITY: Washington  
;; STATE: D.C.  
;; COUNTRY: USA  
;; ZIP: 20005  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentln Release #1.0, Version #1.30  
;;  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/556,355A  
;; FILING DATE: 13-NOV-1995  
;; CLASSIFICATION: 435  
;;  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/424,921  
;; FILING DATE: 19-APR-1995  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/803,627  
;; FILING DATE: 02-DEC-1991  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/779,846  
;; FILING DATE: 21-OCT-1991  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/657,073  
;; FILING DATE: 19-FEB-1991  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/620,568  
;; FILING DATE: 03-DEC-1990  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Barker, M. Paul  
;; REGISTRATION NUMBER: 32,013  
;; REFERENCE/DOCKET NUMBER: 04121.0004-02  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (202)408-4000  
;; TELEFAX: (202)408-4400  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 3499 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;;  
;; US-08-556-355A-2  
;;  
Query Match 79.0%; Score 15.8; DB 2; Length 3499;  
Best Local Similarity 89.5%; Pred. No. 60;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGCGCCAGAGAGATGAT 19  
DB 777 CAAGCGAGAGAGAGATGAT 795  
||||| |||||||||  
US-07-803-627A-2  
Sequence 2, Application US/07803627A  
Patent No. 5948663  
GENERAL INFORMATION:  
APPLICANT: MATTHUR, Eric J.  
TITLE OF INVENTION: Purified Thermostable Pyrococcus  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finegan, Henderson, Farabow, Garrett &  
STREET: 1300 I Street, N.W.  
CITY: Washington

RESULT 13  
US-08-556-355A-2  
Sequence 2, Application US/08556355A  
Patent No. 586395  
GENERAL INFORMATION:  
APPLICANT: MATTHUR, Eric J.  
TITLE OF INVENTION: Purified Thermostable Pyrococcus

STATE: D.C.  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/803,627A  
FILING DATE: 02-DEC-1991  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/776,552  
FILING DATE: 14-OCT-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/657,073  
FILING DATE: 19-FEB-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/620,568  
FILING DATE: 03-DEC-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Barker, W. Paul  
REGISTRATION NUMBER: 32,013  
REFERENCE/DOCKET NUMBER: 04121.0004-00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)408-4400  
TELEFAX: (202)408-4400  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3499 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-07-803-627A-2

Query Match 79.0%; Score 15.8; DB 2; Length 3499;  
Best Local Similarity 89.5%; Pred. No. 60;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGCGCCAGAGAGATGAT 19  
||||| |||||||  
DB 777 CAAGCGAGAGAGATGAT 795

RESULT 15  
US-09-089-879-6  
Sequence 6, Application US/09089879  
Patent No. 611092  
GENERAL INFORMATION:  
APPLICANT: Williamson, Mark  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING DPT111 AND USES THEREOF  
FILE REFERENCE: 0734/078001  
CURRENT APPLICATION NUMBER: US/09/089,879  
CURRENT FILING DATE: 1998-06-03  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 6  
LENGTH: 1542  
TYPE: DNA  
ORGANISM: Mus musculus  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)..(1542)  
OTHER INFORMATION: n = A,T,C or G  
US-09-089-879-6

Query Match 77.0%; Score 15.4; DB 3; Length 1542;  
Best Local Similarity 94.1%; Pred. No. 85;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 AAGCGCAGAGAGATGA 18

|||||  
DB 483 AAGCGCAGAGAGAGA 499

Search completed: May 17, 2003, 03:41:13  
Job time: 30.1788 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 17, 2003, 01:06:03 ; Search time 48.3444 Seconds

(without alignments)  
533.646 Million cell updates/sec

Title: US-10-052-092-15

Perfect score: 20  
Sequence: 1 caagcgccagagagatgatg 20

Scoring table: IDENTITY\_NDC  
Gapop 10.0, Gapext 1.0

Searched: 810007 seqs, 644969091 residues

Total number of hits satisfying chosen parameters: 1620014

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications, NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	20	100.0	20	US-10-052-092-15
2	20	100.0	392	US-10-052-092-4
3	20	100.0	1237	US-10-052-092-3
4	20	100.0	1788	US-10-052-092-24
5	20	100.0	1788	US-09-853-033-1
6	20	100.0	2092	US-10-052-092-6
7	20	100.0	2092	US-09-952-880A-9
8	20	100.0	2106	US-10-052-092-5
9	20	100.0	4963	US-09-281-674-9
10	20	100.0	4963	US-09-892-227-9
11	20	100.0	6450	US-10-067-710-2
12	20	100.0	6450	US-10-081-563-1
13	20	100.0	6450	US-10-052-092-1
14	20	100.0	6450	US-10-052-092-7
15	20	100.0	6610	US-10-052-092-20
16	20	100.0	465237	US-09-933-267A-1
17	16.4	82.0	1978	US-10-052-092-23
18	15.8	79.0	426	US-09-864-761-23769
19	15.8	79.0	534	US-09-864-761-7038

20	15.8	79.0	2454	9	US-10-060-036-176	Sequence 176, App
21	15.8	79.0	2454	10	US-09-962-436-563	Sequence 563, App
22	15.8	79.0	2480	9	US-10-060-036-4548	Sequence 4548, App
23	15.4	77.0	2000	9	US-09-938-842A-2874	Sequence 2874, App
24	15.2	76.0	148	10	US-09-764-877-681	Sequence 681, App
25	15.2	76.0	183	10	US-09-920-300A-988	Sequence 988, App
26	15.2	76.0	183	12	US-10-033-528-988	Sequence 988, App
27	15.2	76.0	210	10	US-09-764-877-3406	Sequence 3406, App
28	15.2	76.0	395	9	US-09-918-995-4369	Sequence 4369, App
29	15.2	76.0	419	9	US-09-918-995-17697	Sequence 17697, A
30	15.2	76.0	427	9	US-09-918-995-17725	Sequence 17725, A
31	15.2	76.0	791	10	US-09-822-849A-578	Sequence 578, App
32	15.2	76.0	1002	10	US-09-770-445-230	Sequence 230, App
33	15.2	76.0	1828	10	US-09-917-800A-1643	Sequence 1643, App
34	15.2	76.0	145831	10	US-09-963-708-79	Sequence 79, App
35	15.2	76.0	145831	10	US-09-954-456-2116	Sequence 2116, App
36	15	75.0	442	9	US-09-796-692-6887	Sequence 6887, App
37	15	75.0	442	9	US-10-040-862-6887	Sequence 6887, App
38	15	75.0	485	9	US-09-796-692-3867	Sequence 3867, App
39	15	75.0	485	9	US-10-040-862-3867	Sequence 3867, App
40	15	75.0	504	9	US-09-796-692-8497	Sequence 8497, App
41	15	75.0	504	9	US-10-040-862-8497	Sequence 8497, App
42	15	75.0	1789	9	US-09-924-340-109	Sequence 109, App
43	15	75.0	1789	9	US-09-924-600A-109	Sequence 109, App
44	14.8	74.0	365	9	US-09-736-457-391	Sequence 391, App
45	14.8	74.0	365	9	US-09-902-941-391	Sequence 391, App

## ALIGNMENTS

RESULT 1  
US-10-052-092-15  
; Sequence 15, Application US/10052092  
; Publication No. US20030027778A1  
; GENERAL INFORMATION:  
; APPLICANT: Fuqua, Suzanne  
; APPLICANT: Allied, D.  
; APPLICANT: Hopp, Torsten A.  
; APPLICANT: O'Connell, Peter  
; TITLE OR INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therap  
; FILE REFERENCE: P02102052  
; CURRENT FILING DATE: 2002-01-18  
; PRIOR APPLICATION NUMBER: US 60/262,990  
; PRIOR FILING DATE: 2001-01-19  
; PRIOR APPLICATION NUMBER: US 60/304,018  
; PRIOR FILING DATE: 2001-07-09  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 15  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Primer  
US-10-052-092-15

Query Match 100.0%; Score 20; DB 9; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.82;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGCGCCAGAGAGATGATG 20  
|||||  
Db 1 CAAGCGCCAGAGAGATGATG 20

RESULT 2  
US-10-052-092-4  
; Sequence 4, Application US/10052092  
; Publication No. US20030027778A1  
; GENERAL INFORMATION:  
; APPLICANT: Fuqua, Suzanne

```
APPLICANT: Allred, D.
APPLICANT: Hopp, Torsten A.
TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therapeuti
FILE REFERENCE: P021020S2
CURRENT APPLICATION NUMBER: US/10/052,092
CURRENT FILING DATE: 2002-01-18
PRIOR APPLICATION NUMBER: US 60/262,990
PRIOR FILING DATE: 2001-01-19
PRIOR APPLICATION NUMBER: US 60/304,018
PRIOR FILING DATE: 2001-07-09
NUMBER OF SEQ ID NOS: 49
SOFTWARE: PatentIn version 3.1
SEQ ID NO 4
LENGTH: 392
TYPE: DNA
ORGANISM: Human
US-10-052-092-4
```

```
Query Match
Best Local Similarity 100.0%; Score 20; DB 9; Length 392;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 CAAGCGCCAGAGAGATGATG 20
Db 28 CAAGCGCCAGAGAGATGATG 47
```

## RESULT 3

```
US-10-052-092-3
Sequence 3, Application US/10052092
Publication No. US20030027778A1
GENERAL INFORMATION:
APPLICANT: Fugua, Suzanne
APPLICANT: Allred, D.
APPLICANT: Hopp, Torsten A.
TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therapeuti
FILE REFERENCE: P021020S2
CURRENT APPLICATION NUMBER: US/10/052,092
CURRENT FILING DATE: 2002-01-18
PRIOR APPLICATION NUMBER: US 60/262,990
PRIOR FILING DATE: 2001-01-19
PRIOR APPLICATION NUMBER: US 60/304,018
PRIOR FILING DATE: 2001-07-09
NUMBER OF SEQ ID NOS: 49
SOFTWARE: PatentIn version 3.1
SEQ ID NO 3
LENGTH: 1237
TYPE: DNA
ORGANISM: Human
US-10-052-092-3
```

```
Query Match
Best Local Similarity 100.0%; Score 20; DB 9; Length 1237;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 CAAGCGCCAGAGAGATGATG 20
Db 390 CAAGCGCCAGAGAGATGATG 409
```

## RESULT 4

```
US-10-052-092-24
Sequence 24, Application US/10052092
Publication No. US20030027778A1
GENERAL INFORMATION:
APPLICANT: Fugua, Suzanne
APPLICANT: Allred, D.
APPLICANT: Hopp, Torsten A.
APPLICANT: O'Connell, Peter
TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therapeuti
FILE REFERENCE: P021020S2
```

```
CURRENT APPLICATION NUMBER: US/10/052,092
CURRENT FILING DATE: 2002-01-18
PRIOR APPLICATION NUMBER: US 60/262,990
PRIOR FILING DATE: 2001-01-19
PRIOR APPLICATION NUMBER: US 60/304,018
PRIOR FILING DATE: 2001-07-09
NUMBER OF SEQ ID NOS: 49
SOFTWARE: PatentIn version 3.1
SEQ ID NO 24
LENGTH: 1788
TYPE: DNA
ORGANISM: Pig
US-10-052-092-24
```

```
Query Match
Best Local Similarity 100.0%; Score 20; DB 9; Length 1788;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 CAAGCGCCAGAGAGATGATG 20
Db 801 CAAGCGCCAGAGAGATGATG 820
```

## RESULT 5

```
US-09-853-033-1
Sequence 1, Application US/09853033
Patent No. US20020100068A1
GENERAL INFORMATION:
APPLICANT: CHAMRON, PIERRE
APPLICANT: METZGER, DANIEL
TITLE OF INVENTION: TRANSGENIC MOUSE FOR TARGETED RECOMBINATION
FILE REFERENCE: 065691/0222
CURRENT APPLICATION NUMBER: US/09/853,033
CURRENT FILING DATE: 2001-05-11
PRIOR APPLICATION NUMBER: FR 00/12570
PRIOR FILING DATE: 2000-10-03
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 1788
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1788)
US-09-853-033-1
```

```
Query Match
Best Local Similarity 100.0%; Score 20; DB 10; Length 1788;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 CAAGCGCCAGAGAGATGATG 20
Db 801 CAAGCGCCAGAGAGATGATG 820
```

## RESULT 6

```
US-10-052-092-6
Sequence 6, Application US/10052092
Publication No. US20030027778A1
GENERAL INFORMATION:
APPLICANT: Fugua, Suzanne
APPLICANT: Allred, D.
APPLICANT: Hopp, Torsten A.
APPLICANT: O'Connell, Peter
TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therape
FILE REFERENCE: P021020S2
CURRENT APPLICATION NUMBER: US/10/052,092
CURRENT FILING DATE: 2002-01-18
PRIOR APPLICATION NUMBER: US 60/262,990
PRIOR FILING DATE: 2001-01-19
PRIOR APPLICATION NUMBER: US 60/304,018
```

PRIOR FILING DATE: 2001-07-09  
 NUMBER OF SEQ ID NOS: 49  
 SOFTWARE: PatentIn version 3.1  
 SEQ ID NO 6  
 LENGTH: 2092  
 TYPE: DNA  
 ORGANISM: Human  
 US-10-052-092-6

Query Match 100.0%; Score 20; DB 9; Length 2092;  
 Best Local Similarity 100.0%; Pred. No. 0.93;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGCCCGAGAGATGATG 20  
 |||  
 Db 1093 CAAGCCCGAGAGATGATG 1112

RESULT 7  
 US-09-952-680A-9  
 Sequence 9, Application US/09952680A  
 Publication No. US20030087239A1  
 GENERAL INFORMATION:  
 APPLICANT: Stanton, Marly  
 APPLICANT: Epstein, David  
 APPLICANT: Hamaguchi, No. US20030087239A1  
 TITLE OF INVENTION: Target Activated Biosensor and Methods of Using Same  
 FILE REFERENCE: 23239-501  
 CURRENT APPLICATION NUMBER: US/09/952,680A  
 CURRENT FILING DATE: 2001-09-13  
 PRIOR APPLICATION NUMBER: 60/232,454  
 PRIOR FILING DATE: 2000-09-13  
 NUMBER OF SEQ ID NOS: 75  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 9  
 LENGTH: 2092  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 US-09-952-680A-9

Query Match 100.0%; Score 20; DB 9; Length 2092;  
 Best Local Similarity 100.0%; Pred. No. 0.93;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGCCCGAGAGATGATG 20  
 |||  
 Db 1093 CAAGCCCGAGAGATGATG 1112

RESULT 8  
 US-10-052-092-5  
 Sequence 5, Application US/10052092  
 Publication No. US20030027778A1  
 GENERAL INFORMATION:  
 APPLICANT: Fugua, Suzanne  
 APPLICANT: Allred, D.  
 APPLICANT: Hopp, Torsten A.  
 APPLICANT: O'Connell, Peter  
 TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therapeutics  
 FILE REFERENCE: P02102052  
 CURRENT APPLICATION NUMBER: US/10/052,092  
 CURRENT FILING DATE: 2002-01-18  
 PRIOR APPLICATION NUMBER: US 60/262,990  
 PRIOR FILING DATE: 2001-01-19  
 PRIOR APPLICATION NUMBER: US 60/304,018  
 PRIOR FILING DATE: 2001-07-09  
 NUMBER OF SEQ ID NOS: 49  
 SOFTWARE: PatentIn version 3.1  
 SEQ ID NO 5  
 LENGTH: 2106  
 TYPE: DNA  
 ORGANISM: Human  
 US-10-052-092-5

Query Match 100.0%; Score 20; DB 9; Length 2106;  
 Best Local Similarity 100.0%; Pred. No. 0.93;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGCCCGAGAGATGATG 20  
 |||  
 Db 801 CAAGCCCGAGAGATGATG 820

RESULT 9  
 US-09-281-674-9  
 Sequence 9, Application US/09281674  
 Patent No. US20020077307A1  
 GENERAL INFORMATION:  
 APPLICANT: Gossen, Manfred  
 APPLICANT: Bujard, Hermann  
 APPLICANT: Salfield, Jochen  
 APPLICANT: Voss, Jeffrey  
 TITLE OF INVENTION: Methods for Regulating Gene Expression  
 NUMBER OF SEQUENCES: 10  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Lahive & Cockfield  
 STREET: 60 State Street, Suite 510  
 CITY: Boston  
 STATE: Massachusetts  
 COUNTRY: USA  
 ZIP: 02109-1875  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: ASCII text  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/281,674  
 FILING DATE: 30-Mar-1999  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/479,306  
 FILING DATE: <Unknown>  
 APPLICATION NUMBER: 08/260,452  
 FILING DATE: 14-JUN-1994  
 APPLICATION NUMBER: 08/076,327  
 FILING DATE: 14-JUN-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Giulio A. DeConti, Jr.  
 REGISTRATION NUMBER: 31,503  
 REFERENCE/DOCKET NUMBER: Brl-013CP3  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 227-7400  
 TELEFAX: (617) 227-5941  
 INFORMATION FOR SEQ ID NO: 9:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 4963 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: circular  
 MOLECULE TYPE: DNA (genomic)  
 ORIGINAL SOURCE:  
 ORGANISM: Human cytomegalovirus  
 IMMEDIATE SOURCE:  
 CLONE: PUHD BGR4  
 SEQUENCE DESCRIPTION: SEQ ID NO: 9:  
 US-09-281-674-9

Query Match 100.0%; Score 20; DB 10; Length 4963;  
 Best Local Similarity 100.0%; Pred. No. 0.95;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGCCCGAGAGATGATG 20  
 |||  
 Db 1266 CAAGCCCGAGAGATGATG 1285



RESULT 10  
US-09-892-227-9  
Sequence 9, Application US/09892227  
Patent No. US20020152487A1  
GENERAL INFORMATION:  
APPLICANT: Gossen, Manfred  
Bujaard, Hermann  
Salzfeld, Jochen  
Voss, Jeffrey  
TITLE OF INVENTION: Animal Transgenic for a Tetracycline-Controlled Transcription  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Iahive & Cockfield  
STREET: 60 State Street, Suite 510  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/892,227  
FILING DATE: 25-Jun-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/163,269  
FILING DATE: <Unknown>  
APPLICATION NUMBER: 08/076,327  
FILING DATE: 14-JUN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Giulio A. DeConti, Jr.  
REGISTRATION NUMBER: 31,503  
REFERENCE/DOCKET NUMBER: BBI-013CP2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4963 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Human cytomegalovirus  
IMMEDIATE SOURCE:  
CLONE: PUHD BGR4  
SEQUENCE DESCRIPTION: SEQ ID NO: 9:  
US-09-892-227-9  
Query Match 100.0%; Score 20; DB 10; Length 4963;  
Best Local Similarity 100.0%; Pred. No. 0.95;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGGCCAGAGAGATGATG 20  
|||||  
Db 1266 CAAGGCCAGAGAGATGATG 1285

RESULT 11  
US-10-096-710-2  
Sequence 2, Application US/10096710  
Patent No. US20020164581A1  
GENERAL INFORMATION:  
APPLICANT: Chang, Chawmshang  
TITLE OF INVENTION: METHODS FOR IDENTIFYING COMPOUNDS THAT  
TITLE OF INVENTION: DO NOT ATTENUATE THE PROTECTIVE EFFECTS OF ESTROGEN  
FILE REFERENCE: 21108.001002

CURRENT APPLICATION NUMBER: US/10/096,710  
CURRENT FILING DATE: 2002-03-12  
PRIOR APPLICATION NUMBER: 60/275,229  
PRIOR FILING DATE: 2001-03-12  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 6450  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence; note =  
OTHER INFORMATION: synthetic construct  
US-10-096-710-2  
Query Match 100.0%; Score 20; DB 9; Length 6450;  
Best Local Similarity 100.0%; Pred. No. 0.96;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGGCCAGAGAGATGATG 20  
|||||  
Db 1161 CAAGGCCAGAGAGATGATG 1180

RESULT 12  
US-10-081-563-1  
Sequence 1, Application US/10081563  
Publication No. US20020187495A1  
GENERAL INFORMATION:  
APPLICANT: Wake Forest University  
APPLICANT: Harrington, David M.  
APPLICANT: Howard, Timothy D.  
APPLICANT: Hawkins, Gregory A.  
TITLE OF INVENTION: GENETIC POLYMORPHISMS OF ESTROGEN RECEPTOR ALPHA ASSOCIATED WITH  
FILE REFERENCE: 9151-15  
CURRENT APPLICATION NUMBER: US/10/081,563  
CURRENT FILING DATE: 2002-02-22  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1  
LENGTH: 6450  
TYPE: DNA  
ORGANISM: Homo sapiens  
PRATYRE:  
NAME/KEY: CDS  
LOCATION: (361)..(2148)  
OTHER INFORMATION:  
PUBLICATION INFORMATION:  
DATABASE ACCESSION NUMBER: X03635  
DATABASE ENTRY DATE: 1993-09-12  
RELEVANT RESIDUES: (1)..(6450)  
US-10-081-563-1  
Query Match 100.0%; Score 20; DB 9; Length 6450;  
Best Local Similarity 100.0%; Pred. No. 0.96;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGGCCAGAGAGATGATG 20  
|||||  
Db 1161 CAAGGCCAGAGAGATGATG 1180

RESULT 13  
US-10-052-092-1  
Sequence 1, Application US/10052092  
Publication No. US20030027778A1  
GENERAL INFORMATION:  
APPLICANT: Fuqua, Suzanne  
APPLICANT: Allied, D.  
APPLICANT: Hopp, Torsten A.  
APPLICANT: O'Connell, Peter

```

; TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therapent
; FILE REFERENCE: P02102US2
; CURRENT APPLICATION NUMBER: US/10/052,092
; CURRENT FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/262,990
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 60/304,018
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 6450
; TYPE: DNA
; ORGANISM: Human
US-10-052-092-1

Query Match
Best Local Similarity 100.0%; Score 20; DB 9; Length 6450;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGGCCGAGAGATGATG 20
    ||||||||||||||||
Db 1161 CAAGGCCGAGAGATGATG 1180

RESULT 14
US-10-052-092-7
; Sequence 7, Application US/10052092
; Publication No. US2003002778A1
; GENERAL INFORMATION:
; APPLICANT: Fugua, Suzanne
; APPLICANT: Allred, D.
; APPLICANT: Hopp, Torsten A.
; APPLICANT: O'Connell, Peter
; TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therapent
; FILE REFERENCE: P02102US2
; CURRENT APPLICATION NUMBER: US/10/052,092
; CURRENT FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/262,990
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 60/304,018
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 6450
; TYPE: DNA
; ORGANISM: Human
US-10-052-092-7

Query Match
Best Local Similarity 100.0%; Score 20; DB 9; Length 6450;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGGCCGAGAGATGATG 20
    ||||||||||||||||
Db 1161 CAAGGCCGAGAGATGATG 1180

RESULT 15
US-10-052-092-20
; Sequence 20, Application US/10052092
; Publication No. US2003002778A1
; GENERAL INFORMATION:
; APPLICANT: Fugua, Suzanne
; APPLICANT: Allred, D.
; APPLICANT: Hopp, Torsten A.
; APPLICANT: O'Connell, Peter
; TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therapent
; FILE REFERENCE: P02102US2
; CURRENT APPLICATION NUMBER: US/10/052,092
; CURRENT FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/262,990
```

```

; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 60/304,018
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 6610
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Vector pCI-n GL1-HEGO
US-10-052-092-20

Query Match
Best Local Similarity 100.0%; Score 20; DB 9; Length 6610;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGGCCGAGAGATGATG 20
    ||||||||||||||||
Db 2685 CAAGGCCGAGAGATGATG 2704
```

Search completed: May 17, 2003, 11:07:21  
Job time : 56.3444 secs

- 1: em\_esta: \*
- 2: em\_estim: \*
- 3: em\_estin: \*
- 4: em\_estim: \*
- 5: em\_estov: \*
- 6: em\_estpl: \*
- 7: em\_estro: \*
- 8: em\_ltc: \*
- 9: gb\_est1: \*
- 10: gb\_est2: \*
- 11: gb\_ltc: \*
- 12: gb\_est3: \*
- 13: gb\_est4: \*
- 14: gb\_est5: \*
- 15: em\_estim: \*
- 16: em\_estom: \*
- 17: gb\_gss: \*
- 18: em\_gss\_hum: \*
- 19: em\_gss\_jay: \*
- 20: em\_gss\_ph: \*
- 21: em\_gss\_vrt: \*
- 22: em\_gss\_fun: \*
- 23: em\_gss\_man: \*
- 24: em\_gss\_mus: \*
- 25: em\_gss\_other: \*
- 26: em\_gss\_pro: \*
- 27: em\_gss\_too: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
c 1	20	100.0	653	9	A1073549	A1075549 ov45e08_x
	18.4	92.0	641	10	AP076636	AP076636 ESTP38874
3	18	90.0	642	12	BP058627	BP058627 TK31nb08_4
4	17.4	87.0	687	10	BS513707	BS513707 G0318476_5
c 5	17.4	87.0	771	17	AQ249357	AQ249357 L10H18-SF
c 6	16.8	84.0	267	10	BE154100	BE154100 PM1-HT034

C	7	16.8	84.0	319	10	BM554950
C	8	16.8	84.0	327	13	BM60807
C	9	16.8	84.0	370	10	BES92363
C	10	16.8	84.0	386	12	BF042197
C	11	16.8	84.0	422	17	BH552807
C	12	16.8	84.0	433	10	BM554964
C	13	16.8	84.0	438	10	BBS48708
C	14	16.8	84.0	439	10	BBS47728
C	15	16.8	84.0	441	12	BF041001
C	16	16.8	84.0	448	10	AW248151
C	17	16.8	84.0	461	13	BM533229
C	18	16.8	84.0	478	12	BG550260
C	19	16.8	84.0	481	13	BI992392
C	20	16.8	84.0	493	9	AI677133
C	21	16.8	84.0	498	10	BE224744
C	22	16.8	84.0	517	17	BE123344
C	23	16.8	84.0	518	10	BE643520
C	24	16.8	84.0	520	10	BE643575
C	25	16.8	84.0	539	10	BE012841
C	26	16.8	84.0	552	14	BM736653
C	27	16.8	84.0	570	14	BQ163168
C	28	16.8	84.0	570	17	TA189311
C	29	16.8	84.0	571	17	TA297003P
C	30	16.8	84.0	574	10	BE234611
C	31	16.8	84.0	629	17	AZ462582
C	32	16.8	84.0	662	12	BF753032
C	33	16.8	84.0	667	13	BF753043
C	34	16.8	84.0	743	17	BH606374
C	35	16.8	84.0	805	17	AZ209248
C	36	16.8	84.0	846	13	BI686571
C	37	16.8	84.0	860	13	BI695840
C	38	16.8	84.0	909	14	BQ714566
C	39	16.8	84.0	943	13	BI546546
C	40	16.8	84.0	1108	12	BG482451
C	41	16.8	84.0	2379	11	AK014006
C	42	16.8	82.0	118	12	BF818483
C	43	16.4	82.0	207	17	AI755595
C	44	16.4	82.0	210	17	AI755593
C	45	16.4	82.0	295	14	L33668
						BM650507 37558 MAR
						AM549070 957042H04
						BES92363 WSI_93_H0
						BF042197 BPF250002A
						BH552807 BOHOV61FTR
						AM549664 37542 MAR
						BBS48708 BBS48708
						BBS47728 BBS47728
						BF041001 BPF250019B
						AW248151 281939.5
						BM533229 512628 MA
						BG550260 947039A06
						BI992392 1020061C1
						AI677133 605051A01
						BE224744 945041F08
						BE123344 945041F08
						BE643520 945041F08
						BE643575 945041F08
						BE012841 122679 MA
						BM736653 952051D06
						BQ163168 952081H06
						AI476637 T. brucei
						AL489381 T. brucei
						BE234611 141774 MA
						AZ462582 1M0263K21
						BF753032 nae18e05.
						BG920943 6028022860
						BH606374 BOGHO24FTR
						AZ209248 SR. 0105 A
						BI686571 6033415577
						BI695840 603347109
						BQ714566 AGNRCODRT
						BI546546 603190160
						BG482451 6025271223
						AK014006 Mus muscu
						BF

## ALIGNMENTS

FEATURES	LOCATION	EST
RESULT 1	AI073549/c	
LOCUS	AI073549	653 bp mRNA, linear EST 06-ANG-1998
DEFINITION	ov45e08.x1 Soares_testis.NET Homo sapiens cDNA clone IMAGE:1640294	
ACCESSION	3, similar to gb:MI2674 ESTROGEN RECEPTOR (HUMAN);, mRNA sequence.	
VERSION	AI073549	
KEYWORDS	AI073549.1 GI:3400193	
SOURCE	EST.	
ORGANISM	human.	
REFERENCE	Homo sapiens	
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
TITLE	1 (bases 1 to 653)	
	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .	
	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),	
	Tumor Gene Index	
	Unpublished (1997)	
COMMENT	Contact: Robert Strausberg, Ph.D.	
	Email: cgaps-r@mail.nih.gov	
	cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo	
	, Ph.D.	
	cDNA Library Arrayed by: Greg Lennon, Ph.D.	
	DNA Sequencing by: Washington University Genome Sequencing Center	
	Clone distribution: NCI-CGAP clone distribution information can be	
	found through the I.M.A.G.E. Consortium/LNLW at:	
	<a href="http://www.bio.lnl.gov/bbrp/image/image.html">www-bio.lnl.gov/bbrp/image/image.html</a>	
	Seq primer: -40m13 fwd. RT from Amersham	
	High quality sequence stop: 483.	
	Location/Qualifiers	

## source

1. 653

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_image="IMAGE:1640294"

/clone\_lib="Soares\_testis\_NHT"

/sex="male"

/lab\_host="DH10B"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker. Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5', TGTTACCATCTGAGTGGAGCGCGCCGCAATTTTCTTTTCTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization to Cols, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 161 a 169 c 142 g 181 t

## ORIGIN

## Query Match

Best Local Similarity 100.0%; Score 20; DB 9; Length 653;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CAAGCCGCGAGAGATGATG 20

DB 612 CAAGCCGCGAGAGATGATG 593

## RESULT 2

LOCUS AM976636

DEFINITION EST388745 MAGE resequences, MAGN Homo sapiens cDNA, mRNA sequence.

ACCESSION AM976636

VERSION AM976636.1 GI:8167867

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 641)

Hegele, P., Qi, R., Aernathy, K., Dharap, S., Gaspar, R., Gay, C., Holt, I.E., Saeed, A.I., Sharov, V., Lee, N.R., Yeatman, T.J., and Quackenbush, J.

Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray

Unpublished (2000)

CONTACT: John Quackenbush

The Institute for Genomic Research

9112 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 3528

Fax: 301 838 0208

Email: johnq@tigr.org

Plate: 363

Seq primer: Forward.

## FEATURES

Location/Qualifiers

1..641

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="MAGE resequences, MAGN"

/note="Vector: pBluescriptPlus"

## BASE COUNT

159 a 175 c 193 g 112 t

## ORIGIN

OY 1 CAAGCCGCGAGAGATGATG 20

DB 597 CAAGCCGCGAGAGATGATG 616

## RESULT 3

LOCUS BF058527

DEFINITION BF058527 x1 NCI-CGAP\_Ov18 Homo sapiens cDNA clone IMAGE:3477350 3' similar to SW-ESR1\_HUMAN P03372 ESTROGEN RECEPTOR 1; mRNA sequence.

ACCESSION BF058527

VERSION BF058527.1 GI:10812423

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 642)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

Unpublished (1997)

CONTACT: Robert Strausberg, Ph.D.

Email: [cgaps@mail.nih.gov](mailto:cgaps@mail.nih.gov)

Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA library preparation: K. Bento Soares, Ph.D. cDNA library arrayed by: Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILNI, send email to: [infoimage.llnl.gov](mailto:infoimage.llnl.gov)

Seq primer: -400P from Gibco

High quality sequence stop: 492.

Location/Qualifiers

## FEATURES

source

1..642

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_image="IMAGE:3477350"

/clone\_lib="NCI-CGAP\_Ov18"

/issue\_type="fibrotheoma"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: ovary; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker. Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCATCTGAGTGGAGCGCGCCGCAATTTTCTTTTCTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

## BASE COUNT

158 a 182 c 196 g 106 t

## ORIGIN

Query Match

Best Local Similarity 90.0%; Score 18; DB 12; Length 642;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CAAGCCGCGAGAGATG 18

DB 609 CAAGCCGCGAGAGATG 626

## RESULT 4

LOCUS BE513707

DEFINITION BE513707 60135475F1 NIH-MGC\_8 Homo sapiens cDNA clone IMAGE:3634170 5', mRNA sequence.

ACCESSION BE513707

VERSION BE513707.1 GI:9720919

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 687)

NIH-MGC <http://mgs.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cga@b-remail.nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
 DNA Library Preparation: Ling Hong/Rubin Laboratory  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNU)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNU at: [image.lnl.gov](http://image.lnl.gov).  
 Plate: ILICM328 row: e column: 19  
 High quality sequence start: 6  
 High quality sequence stop: 441.

## FEATURES

source Location/Qualifiers

1. 687  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="3634170"  
 /clone\_lib="NH\_MGC-8"  
 /tissue\_type="Burkitt lymphoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: lymph. Vector: pDBT; Site.1: XhoI; Site.2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5'  
 adaptor: GGCACGAG(g). Size-selected >500bp for average  
 insert size 1.8kb. Library constructed by Ling Hong in  
 the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."  
 BASE COUNT 184 a 176 c 207 g 120 t  
 ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 687;  
 Best Local Similarity 94.7%; Pred. No. 6e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGCGCCAGAGAGATGATG 20  
 |||||  
 DB 663 AAGCGCCAGAGAGATGAG 681

## RESULT 5

AQ249357/c 771 bp DNA linear GSS 06-OCT-1998  
 LOCUS T10H18-Sp6 TAND Arabidopsis thaliana genomic clone T10H18, DNA  
 DEFINITION

sequence.

ACCESSION AQ249357

VERSION AQ249357.1 GI:3699440

KEYWORDS GSS.

SOURCE

ORGANISM

thale cress.  
 Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 1 (bases 1 to 771)

REFERENCE 1 Peng, J., Dewar, K., Buehler, E., Kim, C., Li, Y., Shinn, P., Sun, H. and  
 Ecker, J.

TITLE BAC End Sequences at ATGC  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Ecker, J.

Arabidopsis thaliana Genome Center  
 University of Pennsylvania  
 Dept. of Biology, University of Pennsylvania, Philadelphia, PA  
 19104  
 Tel: 215-898-9384  
 Fax: 215-898-8780  
 Email: [jecker@atgenom.bio.upenn.edu](mailto:jecker@atgenom.bio.upenn.edu)  
 Seq primer: Sp6  
 Class: BAC ends.

## FEATURES

source Location/Qualifiers

1. 771  
 /organism="Arabidopsis thaliana"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"

/clone="T10H18"  
 /clone\_lib="TAND"  
 /sex="thermaphrodite"  
 /note="Vector: BelosBAC1; Site.1: HindIII; Site.2: HindIII  
 ; Produced by Rod Wing"

BASE COUNT 281 a 154 c 108 g 201 t 27 others  
 ORIGIN

QY 1 CAAGCGCCAGAGAGATGAT 19  
 |||||  
 DB 73 CAAGCGCCAGAGAGATGAT 55

## RESULT 6

BE154100/c 267 bp mRNA linear EST 21-JUN-2000

LOCUS PM1-HT0340-091199-001-808 HT0340 Homo sapiens cDNA, mRNA sequence.

DEFINITION BE154100

ACCESSION BE154100.1 GI:8616821

VERSION

KEYWORDS EST.

SOURCE

ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 267)

REFERENCE Dias Neto, E., Garcia Correa, R., Verjowski-Almeida, S., Briones, M. R.,  
 Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordino, S., Costa, F. F.,  
 Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bala, G. S., Simpson, D. H.,  
 Brunstein, A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare,  
 M. J., Soares, P., Brentani, R. R., Reis, L. F., de Souza, S. J. and  
 Simpson, A. U.

Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 MEDLINE 20202663

## COMMENT

Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?l=et2-PM1-HT0340-091>)  
 199-001-008et3-199-11-09et4-1)

Seq primer: puc 18 forward  
 High quality sequence start: 22  
 High quality sequence stop: 267.

## FEATURES

source Location/Qualifiers

1. 267

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="HT0340"

/dev\_stage="Adult"

/note="Organ: head/neck; Vector: puc18; Site.1: SmaI;  
 Site.2: SmaI; A mini-library was made by cloning products  
 derived from ORSSTES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the puc 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."

BASE COUNT 67 a 63 c 77 g 60 t

## ORIGIN

Query Match 84.0%; Score 16.8; DB 10; Length 267;  
 Best Local Similarity 90.0%; Pred. No. 6.9e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGCGCCAGAGATGATG 20  
 DB 146 CAAGCGCCAGAGATGATG 127

RESULT 7  
 AM354950/c 319 bp mRNA linear EST 25-APR-2001

LOCUS 319 bp mRNA linear EST 25-APR-2001  
 DEFINITION 37558 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.  
 ACCESSION AM354950  
 VERSION AM354950.1 GI:6853940  
 KEYWORDS EST.  
 SOURCE COW  
 ORGANISM Bos taurus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 Bovidae; Bovinae; Bos.  
 1 (bases 1 to 319)  
 Smith, R.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,  
 Casas, E., Wray, J.E., White, J., Cho, J., Fahrnenkung, S.C., Bennett,  
 G.L., Heaton, M.P., Laegreid, W.M., Rohrer, G.A., Chitko-McKown, C.G.,  
 Perlee, G., Holt, L.I., Karamycheva, S., Liang, F., Quackenbush, J. and  
 Keeler, J.W.  
 Sequence evaluation of four pooled-tissue normalized bovine cDNA  
 libraries and construction of a gene index for cattle  
 Genome Res. 11 (4), 626-630 (2001)

TITLE  
 JOURNAL MEDLINE  
 21180013

COMMENT  
 Contact: Smith TPL  
 USDA, ARS, US Meat Animal Research Center  
 PO Box 166, Clay Center, NE 68933-0166, USA  
 Tel: 402 762 4366  
 Fax: 402 762 4390  
 Email: smithemall@marc.usda.gov  
 Single pass sequencing. Bases called and trimmed with phred  
 v0.980904.e. Vector identified by cross-match with the -mmscore 20  
 and -mismatch 12 options.  
 PCR Primers  
 FORWARD: AGGAACAGCATGACATG  
 BACKWARD: GTTTCACATGACGACG  
 Plate: 22 row: J column: 16  
 Seq primer: ATTGAGTGAACATGATG.  
 Location/Qualifiers  
 1. 319  
 /organism="Bos taurus"  
 /db\_xref="taxon:9913"  
 /clone\_lib="MARC 1BOV"  
 /tissue\_type="pooled"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV SPORT6; Site\_1: NotI; Site\_2: SalI;  
 library made from pooled tissue from lymph node, ovary,  
 fat, hypothalamus, and pituitary."

BASE COUNT 79 a 109 g 54 t  
 ORIGIN

Query Match 84.0%; Score 16.8; DB 10; Length 319;  
 Best Local Similarity 90.0%; Pred. No. 7.6e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGCGCCAGAGATGATG 20  
 DB 60 CAAGCGCCAGAGATGATG 41

RESULT 8  
 BM660907/c 327 bp mRNA linear EST 27-FEB-2002

LOCUS 327 bp mRNA linear EST 27-FEB-2002  
 DEFINITION 952042H04.xl 952 - BMS tissue from Walbot Lab (reduced rRNA) Zea  
 mays cDNA, mRNA sequence.  
 ACCESSION BM660907  
 VERSION BM660907.1 GI:18964264  
 KEYWORDS EST.  
 SOURCE Zea mays.

ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
 Clade; Panicoideae; Andropogoneae; Zea.  
 1 (bases 1 to 327)  
 Walbot, V.  
 Maize ESTs from various cDNA libraries sequenced at Stanford  
 University  
 Unpublished (1999)  
 Contact: Walbot V  
 Department of Biological Sciences  
 Stanford University  
 855 California Ave, Palo Alto, CA 94304, USA  
 Tel: 650 723 2227  
 Fax: 650 725 8221  
 Email: walbot@stanford.edu  
 Plate: 952042 row: H column: 04.  
 Location/Qualifiers  
 1. 327  
 /organism="Zea mays"  
 /cultivar="BMS (Black Mexican Sweet)"  
 /db\_xref="taxon:4577"  
 /clone\_lib="952 - BMS tissue from Walbot Lab (reduced rRNA)"  
 /tissue\_type="suspension culture"  
 /dev\_stage="mixed logarithmic and stationary growth  
 phases"  
 /lab\_host="DH10B"  
 /note="Vector: pUC19; Site\_1: EcoRI; Site\_2: EcoRI. The  
 library was prepared by George Rudenko using poly (A)  
 selected RNA and Universal Riboclone cDNA Synthesis System  
 (Promega). cDNA was synthesized using both random and  
 oligo(dT) primers in separate reactions and equipped with  
 EcoRI adaptors. Library was size-fractionated on agarose  
 gels (for insert size >400bp) and non-directionally cloned  
 into EcoRI-digested pUC19 vector. Blue/white selection on  
 carbenicillin-containing plates was used to recover  
 positive clones."

BASE COUNT 86 a 88 c 65 g 86 t 2 others  
 ORIGIN

Query Match 84.0%; Score 16.8; DB 13; Length 327;  
 Best Local Similarity 90.0%; Pred. No. 7.7e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGCGCCAGAGATGATG 20  
 DB 302 CAAGCGCCAGAGATGATG 283

RESULT 9  
 BE592363 370 bp mRNA linear EST 18-AUG-2000

LOCUS 370 bp mRNA linear EST 18-AUG-2000  
 DEFINITION WS1.93.H01.g1.A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,  
 mRNA sequence.  
 ACCESSION BE592363  
 VERSION BE592363.1 GI:9847436  
 KEYWORDS EST.  
 SOURCE sorghum.  
 ORGANISM Sorghum bicolor  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
 Clade; Panicoideae; Andropogoneae; Sorghum.  
 1 (bases 1 to 370)  
 Cordonnier-Pratt, M.-M., Gingle, A., Marasala, C., Sudman, M. and Pratt,  
 L.H.  
 An EST database from Sorghum: water-stressed plants  
 Unpublished (2000)  
 Contact: Cordonnier-Pratt MM  
 Laboratory for Genomics and Bioinformatics  
 The University of Georgia, Department of Plant Biology  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860

Fax: 706 583 0210  
 Email: mmpatt@uga.edu  
 Sequences have been trimmed to exclude polyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.  
 Seq primer: PolYTmX  
 High quality sequence start: 6  
 High quality sequence stop: 352  
 POLYA-No.

## FEATURES

SOURCE

Location/Qualifiers

1..370  
 /organism="Sorghum bicolor"  
 /db\_xref="taxon:4558"  
 /clone\_lib="Water-stressed 1 (WS1)"  
 /note="Organ: Mix of 5-week old plants on days 7 & 8 after water was withheld; Vector: lambda zap; Site\_1: XhoI; Site\_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."

## BASE COUNT

100 a 76 c 103 g 90 t 1 others

## ORIGIN

Query Match 84.0%; Score 16.8; DB 10; Length 370;  
 Best Local Similarity 90.0%; Pred. No. 8.3e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGCGCCGAGAGATGATG 20

Db 40 CAGCGAAGAGAGATGATG 59

## RESULT 10

BF042797/c

LOCUS BF042797 386 bp mRNA linear EST 10-OCT-2000  
 DEFINITION BF250002A10F10 Soares normalized bovine placenta Bos taurus cDNA

ACCESSION BF042797  
 VERSION BF042797  
 KEYWORDS EST.

SOURCE BF042797.1 GI:10759852  
 ORGANISM EST.

## REFERENCE

AUTHORS

Bos taurus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bos.  
 1 (bases 1 to 386)  
 Lewin,H.A., Soares,M.B., Rebeiz,M., Pardinas,T., Liu,L. and Larson,J.H.

## TITLE

BOVINE ESTS

JOURNAL

COMMENT

Unpublished (2000)  
 Contact: Lewin, H. A.  
 W. M. Keck Center for Comparative and Functional Genomics  
 University of Illinois at Urbana-Champaign  
 340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL 61801, USA  
 Tel: 217 333 5998  
 Fax: 217 244 5617  
 Email: h-lewin@uiuc.edu

Funding for cattle EST sequencing was provided by the USDA National Research Initiative, Animal Genome Resource Grant AG 99-3205-8534 to H. A. Lewin and J. B. Womack. Base Calling/Quality Scores: PHRED from Washington University Genome Center. Vector Trimmi g:  
 Cross match from Washington University Genome Center PHRAP suite.  
 This sequence is vector free and at least 200 bp in length. REPEAT IN THE SEQUENCE LOW\_complexity STRAND (+) ELEMENT GC-rich LOCATION [30,81].

PCR PRIMERS  
 FORWARD: TAATAGACTCATATAGG  
 BACKWARD: ATTAACCTCCTAAG

Insert Length: 386 Std Error: 0.00  
 Plate: BP250002A10 row: F column: 10  
 Seq primer: ACGGATACCAATTCACACGA

High quality sequence stop: 386.  
 Location/Qualifiers

## FEATURES

SOURCE

1..386  
 /organism="Bos taurus"  
 /db\_xref="taxon:9913"  
 /clone="BP250002A10F10"  
 /clone\_lib="Soares normalized bovine placenta"  
 /sex="female"  
 /lab\_host="DH10B"  
 /note="Organ: placenta; Vector: pUT3pac; Site\_1: EcoRI; Site\_2: NotI; The cDNA library was contributed by the Soares laboratory and it was constructed and normalized as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome Research 6(9): 791-806."

## BASE COUNT

83 a 103 c 141 g 59 t

## ORIGIN

Query Match 84.0%; Score 16.8; DB 12; Length 386;  
 Best Local Similarity 90.0%; Pred. No. 8.5e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGCGCCGAGAGATGATG 20

Db 220 CGAGCCCGAGAGATGATG 201

## RESULT 11

BH552907/c

LOCUS BH552907 422 bp DNA linear GSS 14-DEC-2001  
 DEFINITION BOHQV61R BOHQ Brassica oleracea genomic clone BOHQV61, DNA sequence.

ACCESSION BH552907  
 VERSION BH552907.1 GI:17804687  
 KEYWORDS GSS.

## SOURCE

ORGANISM

Brassica oleracea.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

## REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Town,C.D., Van Aken,S., Uterback,T. and Fraser,C.M.  
 Whole genome shotgun sequencing of Brassica oleracea  
 unpublished (2001)  
 Other GSSs: BOHQV61R  
 Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.  
 Tel: 301-838-3523  
 Fax: 301-838-0208

Email: cdtown@tigr.org  
 DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TR  
 Class: sheared ends.

## FEATURES

SOURCE

Location/Qualifiers

1..422  
 /organism="Brassica oleracea"  
 /strain="T01000DH3"  
 /db\_xref="taxon:3712"  
 /clone="BOHQV61"  
 /clone\_lib="BOHQ"  
 /note="Vector: pHOSt; Site\_1: BstXI; 2-3 kb sheared genomic DNA inserted into pHOSt using BstXI linkers"

## BASE COUNT

120 a 108 c 94 g 100 t

## ORIGIN

Query Match 84.0%; Score 16.8; DB 17; Length 422;  
 Best Local Similarity 90.0%; Pred. No. 8.3e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGCGCCGAGAGATGATG 20

Db 62 CAAGCGCCGAGAGATGATG 43

## RESULT 12

```

TITLE
JOURNAL
COMMENT
    Unpublished (2001)
    Contact: Yoshihide Hayashizaki
    Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suenho-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan Tel.: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/ Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Komno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) Wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Itaya,M., Ohara,E., Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Komno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) Please visit our web site (http://genome.gsc.riken.go.jp) for further details. e mouse tissues. FEATURES source location/Qualifiers 1..438 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone_id="F930004E21" /clone.lib="RIKEN full-length enriched, adult inner ear" /tissue_type="inner ear" /dev_stage="adult" /note="pooled tissues ; (tissue_type=cerebellum, dev_stage=16 days neonate, sex=mixed) , (tissue_type=cerebellum, dev_stage=0 day neonate, sex=mixed) , (tissue_type=hippocampus, dev_stage=adult, sex=male) , (tissue_type=whole body, dev_stage=9 days embryo, sex=mixed) , (tissue_type=lung, dev_stage=13 days embryo, sex=mixed)" 100 a 119 g 143 g 76 t 20 QY 1 CAAGCGCCAGAGATGATG 20 11111111111111111111 Db 160 CAAAGCGCCAGAGATGCTGT 179 Query Match 84.0% Score 16.8; DB 10; Length 438; Best Local Similarity 90.0%; Pred. No. 9, len+02; Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0; RESULT 14 BB847728 439 bp mRNA linear EST 26-NOV-2001 LOCUS BB847728 RIKEN full-length enriched, adult male kidney Mus musculus DEFINITION BB847728 RIKEN full-length enriched, adult male kidney Mus musculus CDNA clone F530101F07 5', mRNA sequence. ACCESSION BB847728 GI:17086103 VERSION BB847728.1 GI:17086103 KEYWORDS

```





Mon May 19 10:04:46 2003

us-10-052-092-15.rst

Page 8

Job time : 937.993 secs

---